THE UNIVERSITY
of EDINBURGH

The contribution of *Lsh* to DNA methylation reprogramming in embryonic stem cell, epiblast stem cell and embryoid body model systems

Ailsa Clare Revuelta

THE UNIVERSITY *of* EDINBURGH

Presented for the degree of Doctor of Philosophy

The University of Edinburgh

2018

Declaration

I declare that this thesis has been composed by me, and that all the work is my own unless otherwise stated.

This work has not been submitted for any other degree.

Ailsa Revuelta

March 2018

Acknowledgements

I would firstly like to thank my supervisor, Richard, for your guidance, encouragement and positivity throughout my PhD. Secondly, I would like to thank my day-to-day supervisor, Donncha, for your advice, expertise and bioinformatic analysis. The supervision from both of you allowed me to develop as a scientist and to gain confidence over the past four years.

I would also like to express my gratitude to my thesis committee members Andrew Jackson, Ian Adams and Mandy Drake for your time, advice and helpful discussions at my yearly progress meetings. Thank you also to Andy Finch, Jimi Wills and David Hay for optimising the methylation mass spectrometry assay, and for your help with analysing the data. Thanks also to Yatendra and Heidi for your help with EpiSC culture, and to all members of the Genome Regulation section that have given me advice or reagents throughout my PhD.

I would also like to thank past and present members of the Meehan lab for your help, advice and friendship over the past four years. In particular, thank you to lab mama Heidi for the immense amount of cell culture and imaging help you've given me, and for brightening up the lab with your cheerfulness. Thank you also to my booze buddy Katy for letting me moan at you and for keeping me vaguely sane for the past four years – I'll miss our 5.01pm Friday TGI drinks and having someone to lose the plot with at conferences! Thank you to fellow Team Lsh member Leanne for your blastocyst culture and bioinformatics help. Thank you to past lab members Raffy, Ruchi and Michael for your help and useful discussions. Thank you also to past lab member Jen for organising all our social group's activites and outings – we miss you!

Thank you to all my friends at the IGMM that made my time there more fun. Thanks to Fiona for regularly feeding us healthy food, and to Liv and Katy for the gossip sessions that followed over a glass of wine (or disco juice). Thanks also to Louise and Flora for the long lunches and brilliant baking contributions throughout our PhDs.

This thesis would not have been possible if it were not for a number of people outwith the IGMM. Thank you to my parents and sister for your unwavering support and belief

in me throughout my PhD, and throughout my life in general. Most of the credit for me completing this thesis goes to my chef, my copy-editor and my chief motivator, Craig. Thank you for your words of wisdom on the difficult days, and, most impressively, for putting up with me throughout the past four years. Lastly, thanks to my wee dug Barnz for getting me through the hard days with lots of love and mummycuddles.

Abstract

DNA methylation is a key epigenetic mark which undergoes global reprogramming during early mammalian embryonic development, resulting in almost complete erasure of the mark after fertilisation of the zygote. Genome-wide patterns of DNA methylation are subsequently re-established in the implanting blastocyst by *de novo* DNA methyltransferases *Dnmt3a* and *Dnmt3b* along with their catalytically inactive co-factor *Dnmt3l*, while these DNA methylation patterns are maintained through cell divisions by maintenance methyltransferase *Dnmt1*. The exact mechanisms by which these DNA methyltransferase enzymes are targeted to specific genomic regions remain unclear, but may involve interaction with modified histones and/or the participation of co-factors. *Lsh* (lymphoid specific helicase), a putative chromatin remodelling helicase, has been implicated in facilitating *de novo* methylation, as *Lsh* knockout embryos and derived somatic cell lines display substantial but specific DNA methylation losses at repetitive elements and single copy genes.

This study aims to define the requirement for *Lsh* in establishing *de novo* DNA methylation and gene expression patterns during the early stages of mouse embryonic development. The '2i' culture system using two small molecule kinase inhibitors was harnessed to convert *lsh*$^{-/-}$ mouse embryonic stem cells (mESCs) to a hypomethylated 'ground state' of pluripotency. Culture conditions were then altered to transition these ground state mESCs to cells representing later, more methylated stages of development ('serum' mESCs, epiblast stem cells and embryoid bodies). Implementation of this model system suggests that *Lsh* does not contribute to DNA methylation establishment in a pluripotent context, but rather is important for facilitating *de novo* DNA methylation during differentiation to culture models representing later developmental stages. These investigations also reveal that *Lsh* differentially regulates DNA methylation at major and minor satellite repeats depending on cellular context, and that this regulation may involve a role for *Lsh* in maintenance of DNA methylation.

Lay summary

DNA methylation is an essential biological function in which a chemical tag (a methyl group) is added to DNA, affecting many important processes such as how genes are turned off/on and mammalian embryo development. During the early stages after an embryo is fertilised, DNA methylation undergoes dramatic changes during which it is erased then re-established in precise patterns throughout the genome. This process controls what genes are turned on and off in highly tuneable stem cells of the embryo to form specific tissues such as brain, skin and kidney. However, it is not fully understood how DNA methylation alterations are regulated and exactly what proteins are involved.

A protein called LSH is proposed to be involved in re-establishing methylation during mammalian development, as previous gene knockout studies have shown that *Lsh* is important for normal development and for establishment of DNA methylation. The exact mechanism by which *Lsh* contributes to these processes is unknown. During my PhD, I aimed to determine the requirement for *Lsh* in re-establishing DNA methylation patterns in cell culture models equivalent to the early stages of mouse embryonic development. Technology to examine embryo development in real-time remains lacking, therefore I used *in vitro* mouse embryonic stem cells lacking a functional *Lsh* gene to observe the impact of its absence upon DNA methylation and gene expression. I then altered cell growth conditions to mimic the stages of embryo development before and after which DNA methylation re-establishment occurs. This revealed that LSH is required to establish full DNA methylation in stem cells just before they begin to form specific tissues.

This work has contributed to a better understanding of the proteins involved in establishing DNA methylation during the early stages of embryonic development. Since starting my PhD, mutations in *Lsh* have been identified to cause the rare immune disorder ICF syndrome (Immunodeficiency, Centromere instability and Facial anomalies syndrome), a key feature of which is reduced DNA methylation at certain regions in the DNA of affected individuals. Therefore, my research has implications for genetic disease by providing insight into how mutations in *Lsh* may perturb embryonic development and cause ICF syndrome.

Table of contents

Declaration i

Acknowledgements ii

Abstract iv

Lay summary v

Table of contents vi

List of Figures xii

List of Tables xv

List of abbreviations xvi

Chapter 1. Introduction 1

1.1 Epigenetics 1

1.2 Chromatin structure and function 3

1.2.1 Classes of chromatin 4

1.2.2 Heterochromatin formation 5

1.2.3 Bivalent chromatin 6

1.2.4 Nucleosome remodelling 7

1.2.5 Repetitive elements 7

1.3 DNA methylation 9

1.3.1 DNA methylation – the mark 10

1.3.2 DNA de-methylation and 5-hydroxymethylcytosine 11

1.3.3 The roles of DNA methylation 11

1.3.3.1 X-chromosome inactivation 12

1.3.3.2 Genomic imprinting 13

1.3.4 Physiological consequences of DNA methylation perturbation 14

1.3.5 DNA methylation distribution across the mammalian genome 15

1.3.5.1 CpG islands 15

1.3.5.2 DNA methylation at gene bodies and genome-wide 17

1.3.5.3 Non-CpG methylation 18

1.3.6 The DNA methylation machinery 19

1.3.6.1 The maintenance methyltransferase – *Dnmt1* 20

1.3.6.2 The *de novo* methyltransferases – Dnmt3 proteins 21

1.3.6.3 Maintenance vs *de novo* methylation 22

1.3.7 DNA methylation readers 23
1.3.8 DNA methylation co-factors 24
1.3.8.1 *Dnmt3l* 24
1.3.8.2 *Uhrf1* 25
1.3.8.3 *Lsh* 25
1.4 Epigenetic mechanisms in embryonic development 39
1.4.1 Epigenetic dynamics during early embryonic development 40
1.4.2 Pluripotent stem cells as a model for early development 43
1.4.3 Isolation and culture of mouse ESCs 43
1.4.4 The ground state of pluripotency 46
1.4.5 Culture models representing later developmental stages 50
1.5 Thesis aims 51
Chapter 2. Materials and Methods 53
2.1 Mammalian cell culture 53
2.1.1 Cell lines and media 53
2.1.2 mESC serum and 2i culture 54
2.1.3 Transition to epiblast stem cells (EpiSCs) 55
2.1.4 Differentiation to embryoid bodies (EBs) 55
2.1.5 Blastocyst derivation and culture 56
2.1.5.1 Preparation of mitotically inactivated feeders 56
2.1.6 Recovery and cryopreservation of mammalian cell lines 57
2.1.7 Counting and harvesting of mammalian cell lines 57
2.2 DNA preparation, manipulation and analysis 58
2.2.1 DNA extraction from mammalian cell pellets 58
2.2.2 Nucleic acid quantification 59
2.2.3 Agarose gel electrophoresis 59
2.2.4 Restriction digestion of genomic DNA 59
2.2.5 Southern blotting 60
2.2.5.1 Probe preparation 61
2.2.6 Mass spectrometry analysis of 5-methycytosine content 61
2.2.7 Bisulfite sequencing 63
2.3 RNA preparation, manipulation and analysis 64
2.3.1 RNA extraction from mammalian cells 64
2.3.2 cDNA preparation 65
2.3.3 qRT-PCR 65
2.4 Protein preparation and analysis 67

2.4.1 Protein extraction 67
2.4.2 Protein quantification 68
2.4.3 SDS-PAGE 68
2.4.4 Western blotting 69
2.5 Cloning techniques 69
2.5.1 PCR amplification 69
2.5.2 DNA fragment extraction from agarose 70
2.5.3 pGEM T Easy cloning 70
2.5.4 Bacterial transformation 70
2.5.5 Bacterial culture and isolation of plasmid DNA 71
2.5.6 Sanger sequencing 71
2.6 Mammalian cell imaging 72
2.6.1 Immunofluorescence 72
2.6.2 Brightfield imaging 73
2.7 Generation and bioinformatic analysis of next-generation sequencing data 74
2.7.1 Transcripome analysis by RNA-seq 74
2.7.1.1 Library preparation and sequencing 74
2.7.1.2 Mapping and normalisation of RNA-seq data 74
2.7.2 DNA methylome analysis by enhanced reduced representation bisulfite sequencing (ERRBS) 75
2.7.2.1 Mapping and normalisation of ERRBS data 75
Chapter 3. *Lsh* is not required to facilitate establishment of DNA methylation during reversion from the ground state of pluripotency 77
3.1 Introduction 77
3.2 Results 78
3.2.1 Characterisation of *lsh*$^{-/-}$ mESCs 78
3.2.2 Derivation of mESCs from *lsh*$^{-/-}$ mouse blastocysts 79
3.2.3 Optimisation of a culture system to modulate global DNA methylation levels in mESCs 82
3.2.3.1 Experimental strategy 82
3.2.3.2 E14 and *lsh*$^{-/-}$ mESCs change morphologically upon adaptation to 2i and reversion into serum 83
3.2.3.3 *Nanog* and *Essrb* expression becomes more homogeneous upon adaptation to 2i 85
3.2.3.4 E14 and *lsh*$^{-/-}$ mESCs exhibit mRNA expression changes indicative of adaptation to and reversion from 2i 92

3.2.3.5 Global DNA methylation levels fully recover in both E14 and *lsh*$^{-/-}$ mESCs following 2i-associated hypomethylation 97

3.2.4 Investigating the requirement for *Dnmt3l* in re-establishment of DNA methylation during reversion from 2i 101

3.2.4.1 WT and *dnmt3l*$^{-/-}$ mESCs morphologically and transcriptionally transition between serum and 2i 102

3.2.4.2 Re-establishment of DNA methylation following reversion from 2i is impeded in *dnmt3l*$^{-/-}$ mESCs 105

3.3 Discussion 108

3.3.1 Colony morphology and gene expression changes are good indicators of adaptation to and reversion from 2i 109

3.3.2 Differences in colony morphology, gene expression and global DNA methylation levels are interconvertible between serum and 2i culture 110

3.3.3 Serum *lsh*$^{-/-}$ mESCs display some characteristics similar to 2i E14 mESCs 111

3.3.4 Global DNA methylation levels fully recover after reversion into serum in *lsh*$^{-/-}$ but not in *dnmt3l*$^{-/-}$ mESCs 111

Chapter 4. Mapping genome-wide DNA methylation and gene expression changes during transition to and from the pluripotent ground state 114

4.1 Introduction 114

4.2 Results 115

4.2.1 Transcriptome analysis by RNA-seq 115

4.2.1.1 Gene expression profiles of E14 and *lsh*$^{-/-}$ mESCs cluster based on culture conditions rather than genotype 115

4.2.1.2 Gene expression profiles of E14 and *lsh*$^{-/-}$ mESCs cultured in serum and 2i are generally interconvertible 117

4.2.1.3 Serum *lsh*$^{-/-}$ mESCs show similar expression of a subset of genes to E14 2i mESCs 119

4.2.2 DNA methylome analysis by ERRBS 121

4.2.2.1 The DNA methylomes of *lsh*$^{-/-}$ mESCs are distinct from those of E14 mESCs 122

4.2.2.2 Global DNA methylation profiles of E14 and *lsh*$^{-/-}$ mESCs cultured in serum or 2i are interconvertible 124

4.2.2.3 Ground state *lsh*$^{-/-}$ mESCs exhibit moderate changes in DNA methylation compared to 2i E14 mESCs 126

4.3 Discussion 130

4.3.1 Transcriptional and DNA methylation profiles are generally interconvertible between serum and 2i culture 130
4.3.2 Culture conditions have more impact on mESC transcriptomes and DNA methylomes than absence of *Lsh* 131
4.3.3 Lack of *Lsh* results in gene expression and DNA methylation profile perturbations . 132
4.3.4 Technical considerations for DNA methylome analysis by ERRBS 133
Chapter 5. Lsh is required for establishment of DNA methylation in epiblast stem cells and embryoid bodies 135
5.1 Introduction 135
5.2 Results 136
5.2.1 Differentiation of 2i mESCs to EBs 136
5.2.1.1 Experimental strategy 136
5.2.1.2 E14 and *lsh*$^{-/-}$ mESCs adapt morphologically during differentiation to EBs ... 137
5.2.1.3 E14 and *lsh*$^{-/-}$ mESCs transcriptionally upregulate lineage-specific markers following differentiation to EBs 138
5.2.1.4 Global DNA methylation levels are not fully established in *lsh*$^{-/-}$ EBs 144
5.2.1.5 DNA methylation establishment at major and minor satellites is impaired in the absence of Lsh during differentiation to EBs 145
5.2.2 Transition of 2i mESCs to EpiSCs 153
5.2.2.1 Experimental strategy 153
5.2.2.2 E14 and *lsh*$^{-/-}$ 2i mESCs change morphologically during transition to EpiSCs 155
5.2.2.3 E14 and *lsh*$^{-/-}$ mESCs exhibit mRNA expression changes indicative of transition to EpiSCs 156
5.2.2.4 Global DNA methylation levels are reduced in *lsh*$^{-/-}$ 2i mESCs and EpiSCs compared to E14 159
5.2.2.5 DNA methylation establishment at major and minor satellites is impaired in the absence of *Lsh* during transition to EpiSCs 160
5.2.3 Transcriptome profiling of E14 and *lsh*$^{-/-}$ EpiSCs 165
5.2.3.1 The transcriptional profiles of EpiSCs are distinct from mESCs 165
5.2.3.2 Genes involved in developmental processes are mis-expressed in *lsh*$^{-/-}$ EpiSCs 166
5.2.4 Summary of DNA methylation perturbations in the absence of *Lsh* 168
5.2.5 *Lsh* is transcriptionally downregulated during differentiation 169
5.3 Discussion 170

5.3.1 Global and local DNA methylation defects following differentiation to EBs and EpiSCs 171

5.3.2 *Lsh* contributes to de novo DNA methylation during transition to EpiSCs 172

5.3.3 Differential regulation of DNA methylation at satellite repeats is dependent on developmental context 173

5.3.4 Role for *Lsh* in maintenance methylation? 173

5.3.5 *Lsh* influences expression of developmental genes in EpiSCs 174

5.3.6 *Lsh* expression is enriched in pluripotent cell states 174

Chapter 6. General discussion 176

6.1 Technical considerations 179

6.2 *Lsh* contribution to DNA methylation during early development 181

6.3 A role for *Lsh* in maintenance methylation during development? 184

6.4 Influence of *Lsh* on gene expression during development 185

6.5 *Lsh* expression during development 187

6.6 *Lsh* function in DNA methylation and chromatin formation at repetitive elements 188

6.7 Concluding remarks 189

References 190

Appendix 217

List of Figures

Figure 1.1 Waddington's ‘epigenetic landscape’, illustrating the concept of epigenetic restriction of developmental potential. 2
Figure 1.2. Overview of the distribution, organisation and abundance of repetitive elements in the mouse genome. 9
Figure 1.3. DNA methylation mechanism. 10
Figure 1.4. Schematic representation of murine *Lsh* gene 26
Figure 1.5. DNA methylation remodelling during embryonic development 42
Figure 3.1. Confirmation of CRISPR-generated *lsh*$^{-/-}$ mESCs 79
Figure 3.2. Derivation of *lsh*$^{-/-}$ mESCs from mouse blastocysts 81
Figure 3.3. Schematic depicting 2i adaptation and serum reversion experimental design. 83
Figure 3.4. E14 and *lsh*$^{-/-}$ mESCs show similar morphological changes after longer term 2i adaptation and reversion back into serum culture. 84
Figure 3.5. *Nanog* and *Esrrb* expression becomes more homogeneous upon adaptation to 2i and reverts to heterogeneous expression following reversion into serum in both E14 and *lsh*$^{-/-}$ mESCs. 86
Figure 3.6. *Nanog* expression becomes more homogeneous upon adaptation to 2i and reverts to heterogeneous expression following reversion into serum in both E14 and *lsh*$^{-/-}$ mESCs 91
Figure 3.7. Key genes undergo expression changes in E14 and *lsh*$^{-/-}$ mESCs during adaptation to and reversion from 2i. 94
Figure 3.8. DNA methylation at major satellites is reduced upon 2i adaptation and re-established following serum reversion 99
Figure 3.9. Global DNA methylation levels recover in both E14 and *lsh*$^{-/-}$ mESCs following 2i-associated hypomethylation 101
Figure 3.10. Confirmation of WT and *dnmt3l*$^{-/-}$ mESCs 102
Figure 3.11. WT and *dnmt3l*$^{-/-}$ mESCs exhibit a degree of morphological transition between serum and 2i culture 103
Figure 3.12. WT and *dnmt3l*$^{-/-}$ mESCs transition between serum and 2i culture transcriptionally 105

Figure 3.13. Global DNA methylation levels do not recover in *dnmt3l*$^{-/-}$ mESCs following 2i-associated hypomethylation 107
Figure 4.1. Gene expression profiles of E14 and *lsh*$^{-/-}$ mESCs cluster based on culture conditions rather than genotype 116
Figure 4.2. Gene expression profiles of E14 and *lsh*$^{-/-}$ mESCs cultured in serum or 2i are generally interconvertible 118
Figure 4.3. Serum *lsh*$^{-/-}$ mESCs show similar expression of a subset of genes to E14 2i mESCs 120
Figure 4.4. The DNA methylomes of *lsh*$^{-/-}$ mESCs are distinct from those of E14 mESCs 123
Figure 4.5. Global DNA methylation profiles of E14 and *lsh*$^{-/-}$ mESCs cultured in serum or 2i are interconvertible 125
Figure 4.6. A modest number of CpGs exhibit changes in 5-mC status in *serum lsh*$^{-/-}$ mESCs compared to serum E14 mESCs 127
Figure 4.7. A moderate number of CpGs lose and gain methylation in 2i *lsh*$^{-/-}$ mESCs compared to 2i E14 mESCs 129
Figure 5.1. Diagram depicting the experimental workflow of the differentiation of 2i mESCs to EBs in the presence/absence of RA 137
Figure 5.2. E14 and *lsh*$^{-/-}$ mESCs both differentiate to EBs in the presence and absence of RA 138
Figure 5.3. E14 and *lsh*$^{-/-}$ mESCs display expression of lineage-associated markers following differentiation to EBs in the presence/absence of RA 140
Figure 5.4. E14 and *lsh*$^{-/-}$ cells both show transcriptional changes in key genes following differentiation to EBs in the presence/absence of RA 143
Figure 5.5. Global DNA methylation levels are not fully established in *lsh*$^{-/-}$ EBs differentiated in the presence of RA 145
Figure 5.6. DNA methylation is incompletely established at major satellites during differentiation of *lsh*$^{-/-}$ EBs in the presence and absence of RA. 146
Figure 5.7. DNA methylation is incompletely established at minor satellites during differentiation of *lsh*$^{-/-}$ EBs in the presence and absence of RA. 149
Figure 5.8. DNA methylation quantification at major and minor satellites in E14 and *lsh*$^{-/-}$ EBs 151

Figure 5.9. Schematic illustrating the transition of 2i mESCs to EpiSCs following continuous culture in FGF2 and Activin A 155
Figure 5.10. E14 and *lsh*$^{-/-}$ mESCs both exhibit morphological changes during transition to EpiSCs 156
Figure 5.11. E14 and *lsh*$^{-/-}$ cells display transcriptional changes in key genes following transition to EpiSCs 158
Figure 5.12. Global DNA methylation levels are not fully established in *lsh*$^{-/-}$ EpiSCs 160
Figure 5.13. Major and minor satellites are hypomethylated in *lsh*$^{-/-}$ EpiSCs. 162
Figure 5.14. *Lsh* is required for WT levels of DNA methylation at major and minor satellites in EpiSCs..... 164
Figure 5.15. The transcriptional profiles of E14 and *lsh*$^{-/-}$ mESCs and EpiSCs principally cluster based on cellular state. 166
Figure 5.16. *Lsh* influences expression of genes involved in developmental processes in EpiSCs..... 167
Figure 5.17. *Lsh* is transcriptionally downregulated upon differentiation 170
Figure 6.1. Schematic summarising the main findings of this thesis..... 178
Figure 6.2. Off-target mutation in CRISPR-generated *lsh*$^{-/-}$ mESCs 180
Figure A1. Secondary antibody-only staining…….……217
Figure A2. A HELLS mutant human somatic cancer cell line displays substantial global hypomethylation……219

List of Tables

Table 1.1. Phenotypes associated with mutant mouse models of selected DNA methylation mediators and co-factors. 20
Table 2.1. Culture media recipes..... 53
Table 2.2. Specialised culture media recipes. 54
Table 2.3. Details of primers used for bisulfite sequencing..... 64
Table 2.4. Target genes and sequences of primers used for qRT-PCR analysis..... 67
Table 2.5. Details of primers used for sequence mutation checks, CRISPR off-target mutation checks and *Lsh* blastocyst genotyping. 72
Table 2.6. Details of primary antibodies used for immunofluorescence. 73
Table 5.1. Table summarising DNA hypomethylated genomic regions in *lsh*$^{-/-}$ cells compared to the corresponding E14 sample. 169
Table 6.1. Details of genomic regions predicted to be the most likely unintended targets of the *Lsh* sgRNA used to generate *lsh*$^{-/-}$ mESCs. 180

List of abbreviations

5-aza dC - 5-aza-2'-deoxycytidine

5-caC - 5-Carboxylcytosine

5-fC - 5-Formylcytosine

5-hmC - 5'-hydroxymethylcytosine

5-mC - 5'-methylcytosine

ATP - Adenosine triphosphate

AZA - 5'-deoxy-2'-azacytidine

CTCF - CCCTC-binding factor

cDNA - Copy DNA

CGI - CpG island

CpG - cytosine base directly preceding guanosine in DNA

DAPI - 4',6-diamidino-2-phenylindole

DDM1 – Decrease in DNA methylation

DEAH – Asp-Glu-Ala-His, protein domain

DMEM - Dulbecco's Modified Eagle Media

DMSO - Dimethyl sulfoxide

DMR – Differentially methylated region

DNA - Deoxy-ribonucleic acid

DNMT - DNA methyltransferase

EBs – Embryoid bodies

ECCs – Embryonal carcinoma cells

EHMT1/2 - Euchromatic histone-lysine N-methyltransferase 1

EpiSCs – Epiblast stem cells

ERV – Endogenous retrovirus

EGCs – Embryonic germ cells

ESCs - Embryonic stem cells

FITC - Fluorescein isothiocyanate

GOrilla – Gene ontology enrichment analysis and visualisation tool

HCA – Hierarchical clustering analysis

HDAC - Histone deacetylase

HELLS – Helicase, lymphoid specific

HELP - HpaII tiny fragment enrichment by ligation-mediated PCR

HGU - Human genetics unit

HMT – histone methyltransferase

Hox - Homeotic/homeobox

HP1 – Heterochromatin protein 1

HPLC – High performance liquid chromatography

IAP - Intercisternal A particle

ICF - Immunodeficiency, centromere instability and facial abnormalities (a syndrome)

ICM – Inner cell mass

ICR - Imprinting control region

iPSCs – Induced pluripotent stem cells

LAD – Lamin B1 attachment domain

LB - Luria-Bertani

LIF – Leukaemia inhibitory factor

LINE - Long interspersed nuclear elements

Lsh – lymphoid specific helicase

LTR – Long terminal repeat

MAP – Methyl-CpG-binding affinity purification

MBPs – Methyl-CpG binding proteins

MBD - Methyl-CpG binding domain

MeCP - Methyl-CpG-binding protein

MeDIP - Methyl-DNA-immunoprecipitation

MEF- Mouse embryonic fibroblast

mESCs – Mouse embryonic stem cells

MLL - Mixed lineage leukemia

ncRNA - Non-coding RNA

NOMe – Nucleosome occupancy and methylome (sequencing)

NSPCs – Neural stem progenitor cells

PAM – protospacer adjacent motif

PASG – Proliferation-associated SNF2-like gene

PCA – Principal component analysis

PCR - Polymerase chain reaction

PGC - Primordial germ cell

PRC1 - Polycomb repressor complex 1

PRC2 - Polycomb repressor complex 2

PWWP - Refers to amino acid motif (proline, tryptophan, tryptophan, proline)

qRT-PCR – Quantitative reverse transcriptase polymerase chain reaction

RA – Retinoic acid

RING - Really interesting new gene (protein domain)

RIPA - Radioimmunoprecipitation

RNA - Ribo-nucleic acid

RRBS – Reduced representation bisulfite sequencing

SAM - S-adenosyl methionine

SAH - S-Adenosyl homocysteine

SDS - Sodium dodecyl sulfate

Seq - Sequencing (e.g. ChIP-seq)

SET - Su(var)3-9, Enhancer-of-zeste and Trithorax (protein domain)

SETDB1 – SET domain bifurcated 1

sgRNA – single-guide RNA

SINE - Short interspersed nuclear elements

SMARCA - SWI/SNF-related, Matrix-associated, Actin-dependent Regulator Chromatin

S-phase - Synthesis (of DNA) phase

SRA - SET and RING associated domain

SUV39H1/2 - Suppressor of Variegation 3-9 Homolog 1/2

SWI/SNF - SWItch/Sucrose Non Fermentable

TAE - Tris-acetate-EDTA

TBE - Tris-borate-EDTA

TBP – TATA-binding protein

TDG - Thymine-DNA glycosylase

TE - Tris EDTA

TET - Ten-eleven translocation (protein family)

TF - Transcription factor

TSS - Transcription start site

TTD – Tandem tudor domain (protein domain)

TTS - Transcription termination site

TxRd - Texas red

UTR - Untranslated region

WGBS – Whole genome bisulfite sequencing

WT – Wildtype

Xist - X-inactive specific transcript

X^a - Active X chromosome

X^i - Inactive X chromosome

Chapter 1. Introduction

1.1 Epigenetics

The development of a complex, multi-cellular adult organism from a single cell is a highly ordered, reproducible process involving coordination between an array of cellular systems. One of the key cellular systems that is hypothesised to direct developmental decisions during this highly-organised process is the heritable, covalent modification of DNA and histones, now collectively termed as "epigenetics". The embryologist Conrad Waddington introduced the term "epigenetics" in 1942, combining the concepts of "epigenesis" (the development of an organism) and "genetics" (the study of heredity and variation). Waddington defined epigenetics as "the whole complex of developmental processes" connecting genotype to phenotype, and further went on to use the word "epigenotype" to describe the biological networks linking disturbances in genotype to observable abnormal phenotypes (Waddington, 1942). As an embryologist, Waddington's definition of epigenetics was focused around the process of development. The definition of epigenetics has itself developed and been extended over time to give us the generally accepted modern definition that epigenetics is the study of stable changes in gene expression that are not explained by changes in DNA sequence (Bird, 2007). Although epigenetic mechanisms also contribute to genome regulation in a somatic cell context, it could be argued that the most impressive example of epigenetics in action is during embryogenesis – where epigenetic modifications are dramatically remodelled to orchestrate the differentiation of a fertilised zygote into an array of highly-specialised cells that constitute the whole adult organism. This concept is illustrated in Waddington's often-cited model of the "epigenetic landscape" (Figure 1.1 A), which depicts the apparently irreversible fate decisions that a cell is required to make during development. These decisions are underpinned by the role of genes, with complex networks of expression acting to determine the pathway the cell will take during progression through development (Figure 1.1 B).

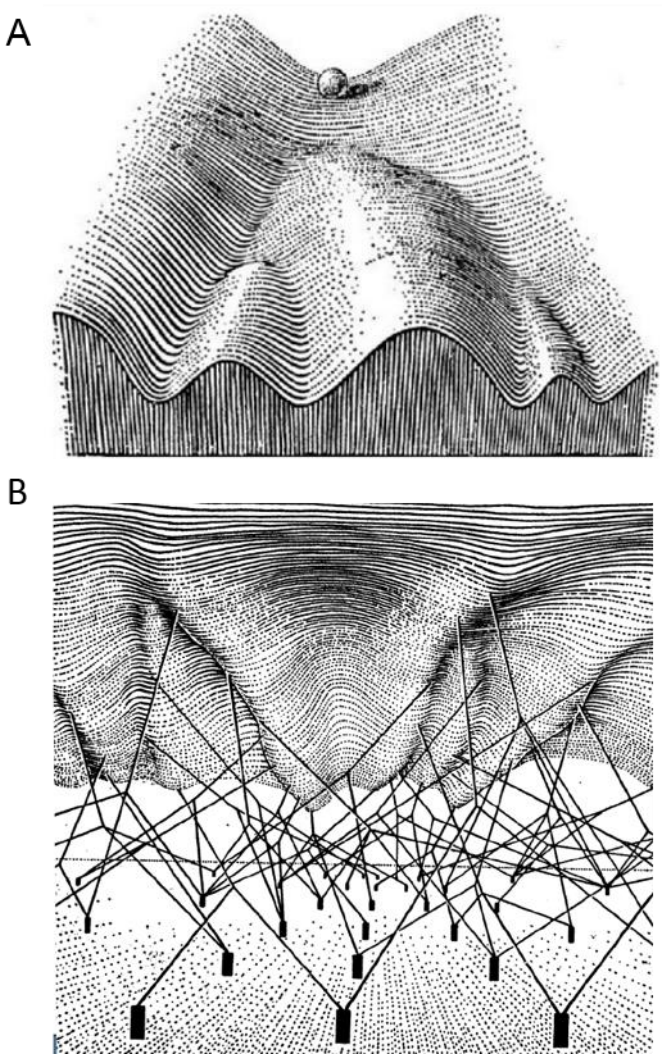


Figure 1.1 Waddington's 'epigenetic landscape', illustrating the concept of epigenetic restriction of developmental potential. A. Cellular development is represented by a ball rolling down a landscape with branching valleys and ridges, signifying the apparently irreversible fate decisions of a developing cell, and the progressive restriction of cellular potential. B. A visual representation of the complex network of underlying genes that act to structure this epigenetic landscape, influencing cell fate decisions and developmental progression. Figures taken from Waddington (1957).

There are now known to be an array of molecular mechanisms involved in epigenetic regulation of the genome. The best studied of these are the direct covalent modification of DNA and of histones, the proteins which DNA are wrapped around to organise and package genetic information in the cell. Research into these two epigenetic mechanisms progressed separately throughout the 1970s and 80s, with the term epigenetics only being adopted to describe research of these modifications in the 1990s, with a rapid increase in research under this term ensuing from the year 2000

(Deichmann, 2016). In recent years, there has also been increasing evidence uncovering the importance of non-coding RNAs (ncRNAs) in epigenetic regulation of gene expression, highlighting the interplay between the functional roles of ncRNAs and the covalent modification of DNA and histones (reviewed by Costa, 2008).

My PhD has involved studying the early stages of development, with a primary focus on the establishment of DNA methylation - the covalent addition of a methyl group to the fifth position of cytosine - and the factors involved in orchestrating its deposition. In particular, I explored the contribution of a DNA methylation co-factor and putative chromatin remodeller *Lsh* (lymphoid specific helicase) to DNA methylation establishment in culture systems representing early embryonic development. This introduction will provide an in-depth discussion of DNA methylation, along with the current knowledge and proposed mechanisms of *Lsh* involvement in its establishment, touching on its potential role in regulation of chromatin structure. I will also summarise the dynamic reprogramming of DNA methylation during the early stages of development, outlining the culture models utilised throughout my PhD to investigate the role of *Lsh* in DNA methylation establishment in a developmental context.

1.2 Chromatin structure and function

The mammalian genome is organised and packaged into chromatin, the structure of which assists regulation of transcription through stable, precise control of a vast array of regulatory elements across the genome. However, the structure of chromatin must retain the ability to undergo dynamic remodelling during periods where extensive changes in gene expression are required, such as those observed in cellular transitions during embryonic development.

The organisation of chromatin is hierarchical, from ‘higher-order’ structures where chromatin is packaged into distinct compact chromosomes, down to lower-level organisation where DNA is wrapped around histone proteins to form ‘nucleosomes’. Nucleosomes are the fundamental repeating units of chromatin and comprise an octamer of the four core histone proteins – H2A, H2B, H3 and H4 – around which 147 base pairs (bp) of DNA are wrapped (Kornberg and Thomas, 1974; Burlingame *et al.*, 1985; Luger *et al.*, 1997). The core histones are predominantly globular, apart from an

unstructured N-terminal ‘tail’ which extends out from the nucleosome (Luger *et al.*, 1997). ‘Linker’ histones – members of the histone H1 family - are structurally distinct from nucleosomal histones and interact with the DNA between nucleosomes and core histone H2A, stabilising both the nucleosome and condensed higher-order chromatin structures (F., Koller and Klug, 1979; reviewed by Robinson and Rhodes, 2006). A fundamental functional feature of histones, and in particular their flexible N-terminal tails, is that they can be adorned with a vast array of post translational modifications, influencing chromatin structure and gene expression (Strahl and Allis, 2000). In terms of epigenetic regulation of transcription, the two main functionally pertinent histone modifications are methylation and acetylation of particular residues, although other modifications such as ubiquitylation also have functional relevance in selected contexts. These histone modifications act to regulate transcription primarily through directly disrupting chromatin contacts or influencing the recruitment of functional non-histone proteins (reviewed by Kouzarides, 2007). Some of the key histone modifications and their contribution to establishment and maintenance of chromatin structure are outlined in the following sections.

1.2.1 Classes of chromatin

Chromatin in mammals can be structurally and functionally partitioned into two categories: ‘euchromatin’, which has a relatively open structure, permissive for high rates of gene expression, and ‘heterochromatin’, which is much more condensed and associated with transcriptional repression. The physically compact state of heterochromatin is reflected by its intensely dark staining with 4,6-diamino-phenylindole (DAPI) and increased resistance to nucleases (Heitz, 1928; Wallrath and Elgin, 1995). This compressed structure reduces the accessibility of DNA to regulatory proteins and transcriptional machinery, resulting in silencing of genes that reside in these heterochromatin regions. Another inherent property of heterochromatin is its ability to spread along chromosomes, demonstrated by the condensation and subsequent silencing of euchromatic transgenes placed in the vicinity of endogenous heterochromatic regions (Wallrath and Elgin, 1995).

Although heterochromatin is primarily associated with transcriptional repression of the genes that it encapsulates, there remains the requirement for dynamic reactivation

of some genes in certain cellular contexts. This is exemplified in the further sub-division of heterochromatin into 'constitutive' and 'facultative' heterochromatin, both with distinct properties and epigenetic signatures. Facultative heterochromatin mostly forms at regions that require dynamic regulation of chromatin structure and transcription in response to developmental and environmental stimuli, including cell-type specific genes and regulatory elements such as enhancers (Brown, 1966; Trojer and Reinberg, 2007). In contrast, the large proportion of repeat-rich sequences in the genome require stable transcriptional silencing across all developmental lineages to prevent their retrotransposition and self-duplication, thereby preserving genome integrity. This is achieved by packaging them into highly condensed inaccessible constitutive heterochromatin, protecting the DNA from contact with the transcriptional machinery or regulatory proteins (Saksouk, Simboeck and Déjardin, 2015).

1.2.2 Heterochromatin formation

Constitutive heterochromatin is characterised by a distinct profile of histone modifications. The most prominent of these is the global absence of histone acetylation (generally associated with actively transcribed euchromatic regions). This is usually accompanied by an enrichment of trimethylation of lysine 9 on histone H3 (H3K9me3), a typical mark of constitutive heterochromatin (Noma, Allis and Grewal, 2001; Martens *et al.*, 2005). This histone modification signature is established through the concerted actions of histone deacetylases (HDACs) and histone methyltransferases (HMTs), resulting in hypoacetylation of histones and hypermethylation of H3K9. In mammals, methylation of H3K9 is catalysed by members of the SET-domain containing family of HMTs (chiefly SET domain bifurcated 1, SETDB1) and related enzymes Suppressor of Variegation 3-9 Homologs 1 and 2 (SUV39H1 and SUV39H2). These enzymes contribute to the majority of H3K9me2 and H3K9me3 (Rea *et al.*, 2000; Schultz *et al.*, 2002), while GLP and G9a (also referred to as euchromatic histone methyltransferase 1 and 2, EHMT1 and EHMT2 respectively) catalyse H3K9me1 and H3K9me2 at heterochromatin (Tachibana *et al.*, 2005). These histone methylation marks can be self-propagated by a mechanism whereby the HMTs bind to existing methylated H3K9 and methylate adjacent nucleosomes, promoting spreading of the mark throughout the region (Zhang *et al.*, 2008). H3K9me2/me3 can

also be bound by the chromodomain of heterochromatin protein 1 (HP1), which can self-oligomerise and form a scaffold to recruit further chromatin modifying proteins including H3K9 methyltransferases and HDACs, forming a feedback loop which further enhances the spread of compacted heterochromatin (Bannister *et al.*, 2001; Canzio *et al.*, 2011). High levels of DNA methylation are a hallmark of this condensed heterochromatin and depend on the presence of the SUV39 and G9a enzymes, highlighting the interplay between the two repressive epigenetic systems (Lehnertz *et al.*, 2003; Epsztejn-Litman *et al.*, 2008). This epigenetically repressed state is stably maintained throughout cell divisions to ensure continued suppression of the underlying repetitive sequences. This involves the incorporation of parental histones harbouring existing modifications into the daughter strands during DNA replication, which in turn stimulates recruitment of H3K9 methyltransferases and re-establishment of the compacted heterochromatic state (Ragunathan, Jih and Moazed, 2014).

The formation of facultative heterochromatin involves a different set of proteins to establish a condensed repressive chromatin state at specific genes in the appropriate developmental context (reviewed by Trojer and Reinberg, 2007). The epigenetic signature that defines facultative heterochromatin is less-well characterised, although the repressive H3K27me3 mark is prominent at the promoters of lineage-specifying genes such as the homeobox (*Hox*) family and correlates with transcriptional silencing. This mark is catalysed by the Polycomb repressive complex 2 (PRC2), which plays a key role in chromatin compaction and heterochromatin assembly at its target genomic loci (Boyer *et al.*, 2006; Margueron *et al.*, 2008; Margueron and Reinberg, 2011). Interestingly, genome-wide H3K27me3 distribution is dependent on global DNA methylation levels (Lynch *et al.*, 2012; Reddington *et al.*, 2013).

1.2.3 Bivalent chromatin

A combination of both ‘active’ and ‘repressive’ histone marks in the same genomic region defines a unique chromatin state referred to as ‘bivalent’ chromatin (Bernstein *et al.*, 2006; reviewed by Voigt, Tee and Reinberg, 2013). Bivalent chromatin contains active H3K4me3 imposed by the SET/mixed lineage leukaemia (MLL) HMT and repressive H3K27me3 catalysed by PRC2. This specific chromatin signature is present primarily at the promoters of developmentally-regulated genes in embryonic stem cells

(ESCs; Mikkelsen *et al.*, 2007). It is thought that these genes that are involved in cell fate decisions are repressed by H3K27me3 in pluripotent cells, but remain 'poised' for timely activation in response to differentiation signals due to the presence of H3K4me3 (Bernstein *et al.*, 2006). This represents another level on which histone modifications contribute to dynamic regulation of chromatin structure and gene expression during development and differentiation.

1.2.4 Nucleosome remodelling

The organisation of the genome into nucleosomes and higher-order chromatin structures is necessary to condense more than 2 metres of DNA into a single nucleus less than 10 microns in length. However, the compaction and association of DNA with many histone and non-histone chromatin components can obstruct access and interpretation of the DNA sequence. The ability to dynamically alter chromatin structure in response to environmental, metabolic and developmental stimuli is critical for quick modification of transcriptional programmes. This alteration of chromatin structure is primarily achieved through the action of a family of adenosine triphosphate (ATP)-dependent chromatin remodelling enzymes. These ATPases can be recruited to specific genomic loci through recognition of histone modifications, allowing targeted regulation of chromatin structure. Once recruited, these chromatin remodellers interrupt histone-DNA interactions in an ATP-dependent manner through disrupting, assembling, exchanging or mobilising nucleosomes (Becker and Hörz, 2002; Clapier and Cairns, 2009). These ATP-dependent chromatin remodelling enzymes are part of large, highly-conserved multi-subunit complexes, of which there are 4 main families of chromatin remodelling complexes in eukaryotes, the best characterised being the switch/sucrose non-fermentable (SWI/SNF) complex first described in yeast (Winston and Carlson, 1992).

1.2.5 Repetitive elements

Sophisticated chromatin organisation mechanisms such as those described above are required to regulate access to the diverse array of sequences present in the mammalian genome. The accumulation of large numbers of repetitive and noncoding sequences contributes to the high complexity of the mammalian genome and the need for multiple

interacting regulatory systems (Franke *et al.*, 2017). In mice in particular, repetitive and noncoding sequences comprise the majority of the genome – 44% repetitive and 52% noncoding – while only around 4% of the genome is accounted for by protein coding genes (Waterston *et al.*, 2002).

The repetitive sequences in the mouse genome are also diverse, ranging from large arrays of repeats associated with centromeres and telomeres, to shorter transposable elements interspersed throughout the genome, sometimes interrupting protein coding genes (outlined in Figure 1.2, Waterston *et al.*, 2002). The role of the clustered arrays of repeat elements in the formation of constitutive heterochromatin at centromeres and telomeres is well-characterised (Almouzni and Probst, 2011). The assembly of heterochromatin at these tandem repeats, referred to as major and minor satellites, is crucial for maintenance of genome stability and repression of transposable elements (Ting *et al.*, 2011; Garcia-Perez, Widmann and Adams, 2016; reviewed by Crichton *et al.*, 2014). Shorter interspersed repeats originate largely from intact retrotransposons, and most can be defined as long or short interspersed nuclear elements (LINEs and SINEs), or endogenous retroviruses (ERVs), a repeat family mainly comprised of intracisternal a particle (IAP) elements which contain long terminal repeats (LTRs; Stocking and Kozak, 2008). Repression of these interspersed repeats is less well defined, although it is equally important to suppress their retrotransposition, which can result in genomic instability and mutations affecting genes and regulatory elements (Burns and Boeke, 2012). However, despite the many detrimental effects of repeat element transposition, there is evidence for its involvement driving genome evolution, as well more nuanced roles in transcriptional regulation (Han, Szak and Boeke, 2004; Cordaux and Batzer, 2009).

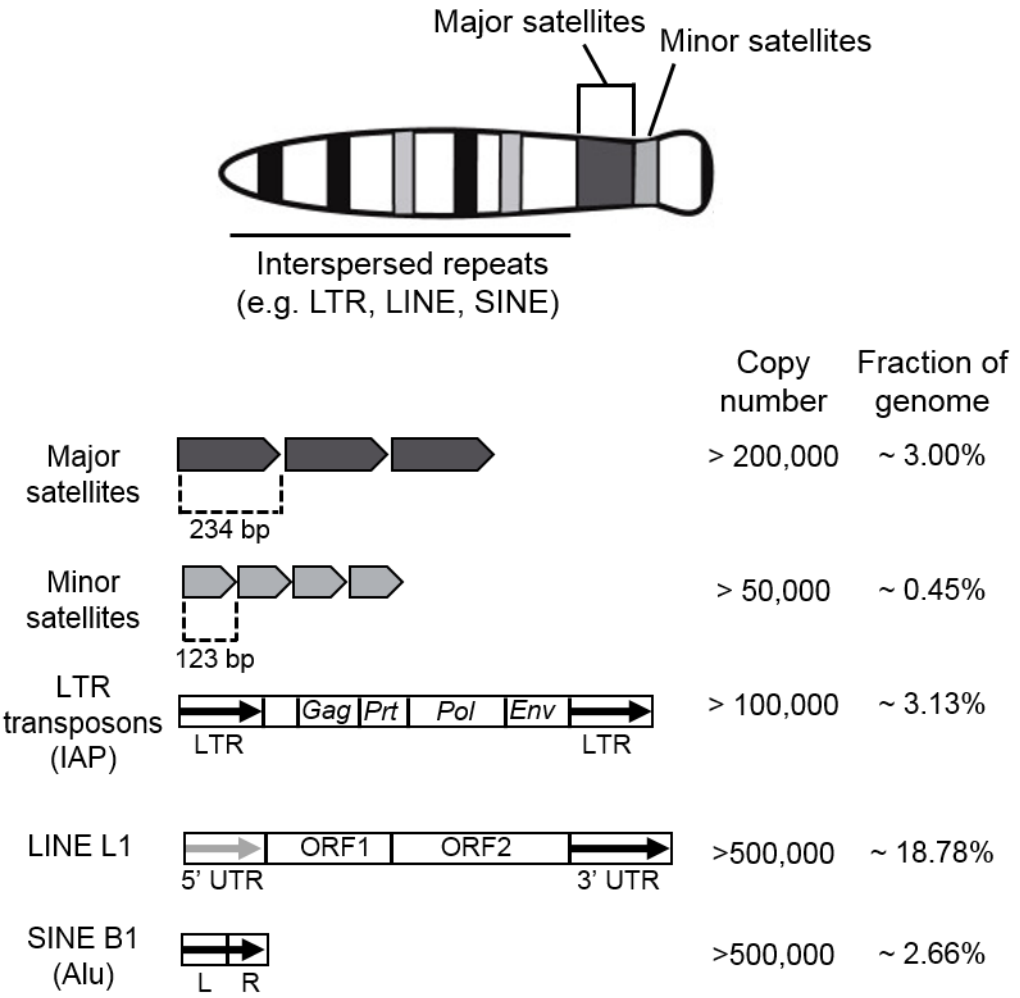


Figure 1.2. Overview of the distribution, organisation and abundance of repetitive elements in the mouse genome. Top panel - the general distribution of repetitive elements is illustrated in the schematic of a mitotic chromosome, highlighting the centromeric and pericentromeric localisation of minor and major satellites, respectively. Lower panels – summary of the main repeat classes demonstrating repeat organisation, copy number and overall abundance in the mouse genome. The numerical values and the concept for this figure was taken from Martens *et al*. (2005).

1.3 DNA methylation

DNA methylation is an epigenetic mechanism involving covalent addition of a methyl group to cytosine, primarily in the context of CpG dinucleotides in the genome. In 1975, this modification was described for the first time as an epigenetic mark that could be established *de novo* or stably inherited through somatic cell divisions, in turn enabling maintenance of gene expression patterns through mitotic divisions (Holliday and Pugh, 1975; Riggs, 1975). These key papers were also the first to associate the mark with transcriptional repression. This association was further established by

studies showing that *in vitro* methylated DNA is transcriptionally inactive following transfection into Xenopus oocytes (Vardimon *et al.*, 1982) or mammalian cell lines (Stein, Razin and Cedar, 1982), functionally linking DNA methylation to transcriptional repression. This contribution to gene silencing was found to be enacted partly through the formation of transcriptionally inactive chromatin, promoting stable maintenance of gene expression signatures (Wolffe and Matzke, 1999). These findings uphold the view that the fundamental role of DNA methylation involves transcriptional repression and heritable maintenance of a silenced chromatin state (Bird, 1984). However, the relationship between DNA methylation and gene silencing has proved difficult to decipher.

1.3.1 DNA methylation – the mark

DNA methylation is the covalent addition of a methyl group to the fifth carbon of the pyrimidine ring of cytosine (5-methylcystosine, 5-mC). This reaction is catalysed by DNA methyltransferase (DNMT) enzymes and requires co-substrate S-adenosyl methionine (SAM) as a methyl group donor (Figure 1.3). 5-mC is a relatively stable mark and is present at similar levels across most adult tissues. However, the levels of DNA methylation can also be dynamically regulated, with the clearest example of this being during the early stages of embryonic development. The molecular mechanisms underlying this dynamic regulation have been further elucidated in recent years due to the discovery of a potential de-methylation pathway involving 5-mC derivatives.

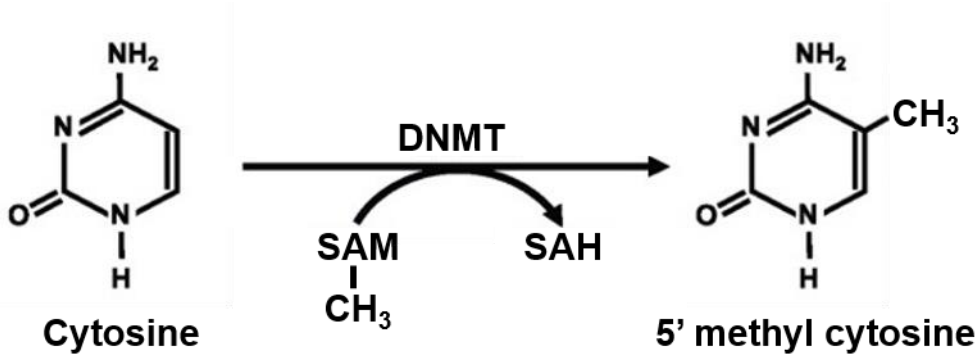


Figure 1.3. DNA methylation mechanism. The covalent addition of a methyl group to the fifth carbon of cytosine is catalysed by DNMT enzymes, using co-substrate SAM as a methyl group donor.

1.3.2 DNA de-methylation and 5-hydroxymethylcytosine

DNA methylation is a potentially reversible epigenetic mechanism. Genomic regions can become de-methylated through a passive process, where 5-mC is diluted due to DNA replication without maintenance methylation. However, there has been much focus on active mechanisms for DNA methylation in recent years due to the rediscovery of 5-hydroxymethylcytosine (5-hmC) and further derivatives 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), which are proposed to represent intermediates of an active de-methylation cycle (Kriaucionis and Heintz, 2009; He *et al.*, 2010; Ito *et al.*, 2011). 5-mC can be oxidised to 5-hmC and further converted to 5-fC and 5-caC by the TET (ten eleven translocation) family of dioxygenases (Ito *et al.*, 2011). These further derivatives can be specifically recognised and excised from the DNA by thymine-DNA glycosylase (TDG) and replaced with an unmodified cytosine, thus completing the de-methylation cycle (He *et al.*, 2010). This process allows dynamic regulation of DNA methylation without the need for DNA replication.

As well as being an intermediate in a de-methylation pathway, 5-hmC is proposed to be a relatively abundant mark with epigenetic functions in its own right. Unlike the fairly uniform levels of 5-mC across tissue and cell types, 5-hmC levels have been shown to vary with tissue type and are most abundant in neuronal tissues (Globisch *et al.*, 2010; Nestor *et al.*, 2012). Furthermore, 5-hmC has been found to be present at gene bodies, where it correlates with levels of transcription (Nestor *et al.*, 2012). These findings link 5-hmC (and/or the TET enzymes) with regulation of transcription, perhaps in particular cellular contexts where the mark is most abundant.

1.3.3 The roles of DNA methylation

The fundamental function of DNA methylation is regarded as promoting stable epigenetic repression of transcription, although the generality of this function across all biological processes is under question. Indeed, genome-wide analysis of promoter methylation implies that only a small subset of genes are transcriptionally regulated by DNA methylation (Weber *et al.*, 2007; Hackett *et al.*, 2012; reviewed by Weber and Schübeler, 2007 and Reddington, Pennings and Meehan, 2013). However, the

crucial role of DNA methylation in gene expression regulation is clearly demonstrated in a couple of key biological contexts. These include the processes of X-chromosome inactivation and imprinting, where the repressive action of DNA methylation is demonstrated most strongly and on a large scale, employing similar mechanisms to ensure appropriate dosage compensation is established during early embryonic development.

1.3.3.1 X-chromosome inactivation

X-chromosome inactivation is a dosage compensation mechanism in which one of the two X-chromosomes present in female mammalian cells is transcriptionally silenced, forming an inactive X-chromosome (X^i). In the early stages of embryogenesis, the paternal X-chromosome is preferentially silenced then reactivated in the epiblast of the blastocyst, where one X-chromsome is then randomly chosen for inactivation (Mak *et al.*, 2004). This inactive X^i established in the cells of the epiblast, is faithfully maintained in all the somatic lineages that these cells generate and throughout the lifespan of the adult organism.

The repression of the X^i is mediated by the expression of the functional long non-coding RNA (lncRNA) *Xist*. *Xist* RNA triggers transcriptional silencing in *cis* by coating the chromosome from which it's expressed, inducing the acquisition of repressive histone marks and the subsequent formation of a silent chromatin state (Clemson *et al.*, 1996; Kohlmaier *et al.*, 2004). Heterochromatin assembly at the X^i is followed by changes in DNA methylation patterns across both X-chromosomes (Hellman and Chess, 2007; Sharp *et al.*, 2011). DNA methylation is established at many promoters of genes that are silenced on the X^i. Depletion of promoter methylation experimentally through 5-aza-2'-deoxycytidine (5-aza dC) treatment or *Dnmt1* conditional deletion results in transcriptional reactivation of these genes and decondensation of the X^i over time (Sado *et al.*, 2000; Csankovszki, Nagy and Jaenisch, 2001). This indicates that DNA methylation is required for stable maintenance of gene repression and the heterochromatic state on the X^i. On the active X-chromosome (X^a), global DNA methylation levels are twice as those of the X^i and show a very different distribution of promoter hypomethylation coupled with extensive gene body

methylation, which is thought to support expression of active genes on the X^a (Hellman and Chess, 2007; Sharp *et al.*, 2011). Finally, DNA methylation contributes to the regulation of *Xist* expression itself, as *de novo* methylation of the *Xist* promoter acts to repress aberrant *Xist* expression from the second X-chromosome (Sado *et al.*, 2004).

1.3.3.2 Genomic imprinting

Genomic imprinting is an epigenetically regulated process that results in the mono-allelic expression of genes in a parental-origin-specific manner. This mono-allelic expression relies on DNA methylation to mark, or 'imprint', either the maternal or paternal allele so that the chromosome homologues can be distinguished from one another. These differentially methylated regions (DMRs) at imprinted genes are established during gametogenesis in a sex-specific manner and are maintained in the resulting offspring (reviewed by Ferguson-Smith, 2011).

Transcriptional regulation of imprinted genes by DNA methylation is often complex and varies depending on the genomic loci and cellular context. The expression of a subset of imprinted genes can be directly repressed by methylation at DMRs that overlap with the gene promoter. However, many imprinted genes are regulated through methylation of a genomic region called the imprinting control region (ICR), which can control expression of a gene or a cluster of genes located more than a megabase away (Lin *et al.*, 2003; Williamson *et al.*, 2006). The regulation of the *Igf2-H19* locus is a well-studied example of how differential ICR methylation can affect the mono-allelic expression of each gene. The unmethylated ICR on the maternal allele binds chromatin insulator CCCTC-binding factor (CTCF), which in turn blocks the interaction of *Igf2* with a shared set of distal enhancers, effectively silencing *Igf2* while *H19* expression is permissible. Reciprocally, CTCF cannot bind to the methylated ICR on the paternal allele, allowing expression of *Igf2* through interaction with the distal enhancers, while *H19* expression is blocked (Bell and Felsenfeld, 2000; Hark *et al.*, 2000). Another transcriptional silencing mechanism involving DNA methylation of ICRs is the regulation of functional lncRNA expression, whose promoters often overlap with ICRs. When expressed, these lncRNAs impress a repressive chromatin state in *cis* that

can encompass and silence multiple imprinted genes through acquisition of histone modifications and histone modifying enzymes (Nagano *et al.*, 2008).

These examples of X-inactivation and imprinting both highlight common mechanisms employed to initiate and maintain long-term mono-allelic silencing through the interplay of DNA methylation, chromatin structure and lncRNA expression.

1.3.4 Physiological consequences of DNA methylation perturbation

The selective expression and repression of genes in a developmental stage- and cellular context-dependent manner is essential for the correct development and function of an organism. The gene mis-expression that is apparent in many human diseases and cancer has been linked to perturbations in DNA methylation patterns globally and in the vicinity of the mis-expressed genes.

The importance of the DNA methylation machinery in normal development and function is highlighted by the severity of the mouse mutant phenotypes, such as the embryonic lethality of *Dnmt1* null mice (Table 1.1). In *Dnmt1* null ESCs, which show a 70% reduction in normal cytosine methylation levels, the rate of mutations that involve gene rearrangements is increased ten-fold (Chen *et al.*, 1998). There is also a massive upregulation of IAP elements in *Dnmt1*$^{-/-}$ embryos and somatic cells, which has been proposed to play a part in the increased genomic instability (Walsh, Chaillet and Bestor, 1998; Dunican *et al.*, 2013). Missense mutations in *de novo* methyltransferase *Dnmt3b* are also associated with genome instability as they are reported to cause immunodeficiency-centromeric instability-facial anomalies (ICF) syndrome, a rare autosomal recessive disorder characterised by chromosomal rearrangements due to hypomethylation of pericentromeric repeats (Hansen *et al.*, 1999; Xu *et al.*, 1999; Ehrlich *et al.*, 2008).

Hypomethylation of transposable and repetitive elements has also been linked to genome instability in the context of cancer, and is known to contribute to cell transformation and cancer progression (Shukla *et al.*, 2013). The general hypomethylation that is observed throughout the genome in cancer is accompanied by the striking hypermethylation of many normally unmethylated CpG islands (CGIs).

This occurs largely at the promoters of tumour suppressor genes as a result of growth selection and acts to promote tumorigenesis (reviewed by Baylin and Jones, 2011 and Sproul and Meehan, 2013). However, this view has been challenged by the observation that most of the genes associated with affected CGIs are already silenced prior to aberrant hypermethylation during cancer development, de-linking transcriptional silencing from aberrant methylation (Sproul *et al.*, 2011, 2012). While mutations in the DNA methylation machinery resulting in aberrant DNA methylation patterns have been linked with acute myeoloid leukaemia and colorectal cancer, aberrant epigenetic programs in cancer are more likely a consequence of genetic changes (Kanai *et al.*, 2003; Ley *et al.*, 2010; Yan *et al.*, 2011).

The disruption of normal patterns of DNA methylation or its writers are associated with many more diseases and disorders, as well as cancer and genomic instability. This emphasises the critical role of this epigenetic mechanism in multiple physiological processes (reviewed by Timp and Feinberg, 2013; Feinberg, Koldobskiy and Göndör, 2016).

1.3.5 DNA methylation distribution across the mammalian genome

Examination of the distribution of DNA methylation throughout the genome can provide insight and understanding into the function of DNA methylation. Techniques involving next-generation sequencing to study genome-wide methylation have rapidly improved over the past decade, allowing more extensive, in-depth and high-resolution characterisation of the methylome than ever before. This has been harnessed to study genome-wide methylation patterns across the genomes of a variety of cell types and organisms, providing further insight into DNA methylation function at various classes of elements throughout the genome.

1.3.5.1 CpG islands

DNA methylation in the mammalian genome is widespread, with generally between 60 and 80% of CpGs being methylated, representing around 4% of all cytosines in the genome (Gruenbaum *et al.*, 1981). This widespread methylation is not uniformly distributed throughout the genome, but rather exhibits a bimodal pattern of distribution. Almost all the CpGs in the mammalian genome (which is itself globally

depleted of CpGs) are methylated. However, there is an enrichment of largely methylation-resistant CpGs in genomic regions referred to as CpG islands (CGIs). These short (approximately 1 kilobases) CG-dense sequences overlap with the promoter regions of more than half the genes in the mammalian genome ('promoter CGIs'), and are particularly prevalent at genes involved in housekeeping and developmental regulation (Larsen *et al.*, 1992; Deaton and Bird, 2011). There are also a large proportion of CGIs that occur in intergenic and intragenic regions not associated with a promoter ('orphan CGIs'). These orphan CGIs share epigenetic features with promoter CGIs but are found to be more frequently methylated during development (Illingworth *et al.*, 2010).

The preservation of the CpG-rich status of CGIs in contrast to the rest of the genome can be accounted for by the constitutive hypomethylation of these regions, as methylated cytosine can be spontaneously or enzymatically converted to thymine by deamination, resulting in CpG loss throughout the genome (Deaton and Bird, 2011). The maintenance of CGI hypomethylation is thought to be partly accounted for by the constitutive methylation of CGIs at H3K4 (predominantly H3K4me2), acting to prevent the association of DNA methyltransferases with CGIs (Weber *et al.*, 2007; Meissner *et al.*, 2008). This is demonstrated by the action of DNA methyltransferase co-factor *Dnmt3l*, which binds unmodified H3K4 only, initiating *de novo* DNA methylation at its intended targets *via Dnmt3a* (Jia *et al.*, 2007; Ooi *et al.*, 2007).

Although most promoter-associated CGIs are hypomethylated in normal tissues, a small subset are methylated (< 3%), which correlates with transcriptional repression of the gene (Illingworth *et al.*, 2008; Maunakea *et al.*, 2010; Hackett *et al.*, 2012). The repressive function of DNA methylation at promoter CGIs is most powerfully demonstrated during the processes of X-inactivation and genomic imprinting, where their methylation results in robust and long-term silencing of the associated gene (Li, Beard and Jaenisch, 1993; Augui, Nora and Heard, 2011). Conversely, the consequences of aberrant CGI promoter methylation are clearly demonstrated in the context of cancer, where CGI hypermethylation in combination with gene body hypomethylation result in the transcriptional silencing of critical growth regulators such as tumour suppressor genes.

1.3.5.2 DNA methylation at gene bodies and genome-wide

DNA methylation patterns beyond the well-studied promoters are less well-characterised, but genome-wide studies have uncovered some intriguing insights into potential links between DNA methylation and transcription in the past few years. In contrast to the methylation-resistant CpG-rich CGIs described previously, the remainder of the mammalian genome is CpG-poor, contains multiple repetitive and transposable elements and exhibits high levels of DNA methylation. Most gene bodies exhibit these features and display extensive methylation. Unlike in CGI promoter regions however, this methylation is not associated with repression. Far from it, gene body methylation has been shown to be a feature of transcriptionally active genes (Wolf *et al.*, 1984). This was elegantly demonstrated in a study that comparatively mapped DNA methylation along the active and inactive X-chromosomes in human cells, correlating increased DNA methylation at gene bodies on the more transcriptionally active X^a while gene body methylation on the transcriptionally inactive X^i was comparatively depleted (Hellman and Chess, 2007). Furthermore, a recent study has provided further insight into the relationship between gene body methylation and transcriptionally active genes, showing that Dnmt3b methylates actively transcribed regions of the genome via its interaction with H3K36me3 (Baubec *et al.*, 2015).

One of the primary proposed functions of hypermethylation at gene bodies is that it acts to prevent inappropriate transcription initiation outwith the canonical TSS, whilst allowing transcriptional elongation to continue undisturbed (Yoder, Walsh and Bestor, 1997; Neri *et al.*, 2017; reviewed by Jones, 2012). This potential role was extended to propose a more general function of DNA methylation globally, acting as a mechanism to block aberrant initiation of transcription at repetitive regions such as retroviruses and LINE1 elements (Yoder, Walsh and Bestor, 1997). The concept of suppression of transcriptional activation by gene body methylation was exemplified by Maunakea *et al.* (2010). This study demonstrated tissue-specific methylation of an intragenic CGI acting as an alternative promoter at the *Shank3* locus in mouse and human cells, resulting in tissue- and cell type-specific expression of alternative transcripts. Other potential roles for DNA methylation at gene bodies have also come to light more

recently, including the link between gene body methylation and regulation of splicing. Genome-wide high resolution mapping of a variety of human cell lines showed that exons are more highly methylated than introns, with a sharp transition at the intron-exon boundary, suggesting a function for 5-mC in splicing (Laurent *et al.*, 2010). A further study demonstrated the effect of gene body DNA methylation on CTCF binding, highlighting the consequences for RNA polymerase II pausing and subsequent inclusion of alternatively spliced exons (Shukla *et al.*, 2011). These studies highlight a previously uncharacterised role for DNA methylation in transcription.

1.3.5.3 Non-CpG methylation

Although methylated cytosine is predominantly present in the context of CpG dinucleotides, it is not exclusively found in this context. The existence of 5-mC in CpT, CpC and particulary CpA contexts has been documented in recent years. It has been shown that nearly one quarter of methylated cytosines in human ESCs were present in a non-CpG context, whereas nearly all DNA methylation in somatic cells occurred within CpG dinucleotides (Lister *et al.*, 2009). This non-CpG methylation disappeared following differentiation of these ESCs, and subsequently reappeared upon restoration of pluripotency in induced pluripotent stem cells (iPSCs; Lister *et al.*, 2009; Ziller *et al.*, 2011). This implies the use of alternative methylation mechanisms in pluripotent cells for regulation of gene expression. However, knockdown of *Dnmt3a* and *Dnmt3b* resulted in a global reduction of non-CpG methylation in human ESCs, showing that the same DNMTs are responsible for establishment of 5-mC independent of the context (Ramsahoye *et al.*, 2000; Ziller *et al.*, 2011). Indeed, high levels of non-CpG methylation in ESCs may be a by-product of high *de novo* methyltransferase activity in these cells (Lei *et al.*, 1996). Like CpG methylation, non-CpG methylation was shown to be enriched in exons of gene bodies compared to introns and depleted in protein-binding sites, promoters and enhancers. Additionally, gene ontology analysis revealed that gene bodies marked by non-CpG methylation were enriched for functions in RNA processing and splicing, suggesting a potentially similar role for CpG methylation present in these regions (Lister *et al.*, 2009).

1.3.6 The DNA methylation machinery

A mechanism for inheritance of DNA methylation patterns through somatic cell divisions was proposed more than 40 years ago in two key publications predicting the existence of an enzyme-mediated mechanism for *de novo* methylation and subsequent maintenance of methylation patterns through recognition of hemimethylated DNA (Holliday, R. & Pugh, 1975; Riggs, 1975). These papers led to the concept of 'maintenance' enzymes that could propagate DNA methylation patterns through generations of somatic cells, and '*de novo*' enzymes that establish these methylation patterns (Holliday & Pugh, 1975). Three conserved enzymes, collectively termed 'DNA methyltransfersases' (DNMTs), have since been identified and placed into these two functional classes. The proposed maintenance methyltransferase, *Dnmt1*, was identified in 1988 followed by the *de novo* methyltransferases *Dnmt3a* and *Dnmt3b* a decade later (Bestor *et al.*, 1988; Masaki Okano, Xie and Li, 1998; Okano *et al.*, 1999). All three have been shown to be essential for normal development (Table 1.1, Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999). A fourth methyltransferase, *Dnmt2*, has also been identified, and rather than contributing to DNA methylation patterns, it has been shown to act as a transfer RNA (tRNA) cytosine methyltransferase with possible roles in the regulation of protein translation (Goll *et al.*, 2006; reviewed by Jeltsch *et al.*, 2017). Furthermore, a new *de novo* methyltransferase, *Dnmt3c*, was identified just last year and appears to play a specific role in methylation of evolutionarily young transposons in the male germ line (Barau *et al.*, 2016). The key features and functions of the main components of the DNA methylation machinery will be discussed in this section along with their contribution to genomic DNA methylation.

Protein	Primary function	Mouse model phenotype	Reference(s)
DNA methyltransferases			
Dnmt1	Maintenance methylation	Embryonic lethal by E8.5; global hypomethylation; loss of gene repression associated with imprinting and X-inactivation; transposon activation, ESCs viable but apoptose upon differentiation	Li *et al.* (1992); Lei *et al.* (1996)
Dnmt3a	*De novo* methylation	Viable but runted; die postnatally (~4weeks); spermatogenesis defects; impaired *de novo* methylation particularly at imprints	Okano *et al.* (1999)
Dnmt3b	*De novo* methylation	Embryonic lethality around E14.5; defective *de novo* methylation at minor satellites	Okano *et al.* (1999)
Dnmt3c	*De novo* methylation	Viable; hypogonadism; male sterility; impaired *de novo* methylation at evolutionary young retrotransposons	Barau *et al.* (2016)
DNA methylation co-factors			
Dnmt3l	*De novo* methylation	Viable; spermatogenesis defects, impaired *de novo* methylation at imprints and retrotransposons in germ cells	Bourc'his et al. (2001); Hata *et al.* (2002)
Uhrf1	Maintenance methylation	Early embryonic lethality; developmental arrest shortly after gastrulation; similar to *Dnmt1* mouse model	Sharif *et al.* (2007)
Lsh	*De novo* methylation(?)	Peri- or post- natal lethality (a few hours to a few weeks after birth); runted; renal abnormalities; skeletal defects; premature ageing phenotype; abnormal meiosis in gametes	Geiman *et al.* (2001); Sun *et al.* (2004)

Table 1.1. Phenotypes associated with mutant mouse models of selected DNA methylation mediators and co-factors.

1.3.6.1 The maintenance methyltransferase – Dnmt1

The ability for DNA methylation patterns to be stably maintained through generations of somatic cells is key to the concept of epigenetic memory, where inheritance of epigenetic patterns propagate transcriptional states through cell divisions. This requires a mechanism whereby DNA methylation patterns can be copied from the

parent strand to the nascent strand during DNA replication in S phase of the cell cycle. The first eukaryotic cytosine methyltransferase to be identified, *Dnmt1*, was shown to have a preference for hemimethylated DNA and to interact with components of the replication machinery during S phase (Bestor *et al.*, 1988; Chuang *et al.*, 1997; Goyal, Reinhardt and Jeltsch, 2006; Schermelleh *et al.*, 2007). Therefore, it is postulated to be a maintenance methylatransferase, acting to preserve patterns of DNA methylation at each replication cycle.

Dnmt1 is the most abundant DNA methyltransferase and is widely expressed in somatic tissues throughout mammalian development, exhibiting particularly high expression levels in dividing cells (Walsh and Bestor, 1999; Goll and Bestor, 2005). *Dnmt1* is essential for normal embryonic development, as *Dnmt1*-targeted knockout mouse models show embryonic lethality around embryonic day (E)8.5 at the onset of gastrulation (Li, Bestor and Jaenisch, 1992; Lei *et al.*, 1996). There is an array of reported defects with these embryos (Table 1.1), the most notable being a dramatic reduction in global DNA methylation levels, resulting in mis-expression of genes associated with imprinting and X-inactivation, as well as elevated IAP element transcription (Li, Beard and Jaenisch, 1993; Lei *et al.*, 1996; Walsh, Chaillet and Bestor, 1998). Despite the severe embryonic lethal phenotype exhibited by *Dnmt1* deficient mice, ESCs lacking *Dnmt1* are viable and divide normally in an undifferentiated state, despite exhibiting very low levels of genomic DNA methylation. However when these ESCs are induced to differentiate they undergo p53-mediated apoptosis (Li, Bestor and Jaenisch, 1992; Lei *et al.*, 1996). This highlights that the severe DNA hypomethylation present in *Dnmt1* null ESCs does not affect cell viability before differentiation, but upon initiation of differentiation *Dnmt1* activity becomes an essential requirement.

1.3.6.2 The de novo methyltransferases – Dnmt3 proteins

Two further members of the DNMT family, *Dnmt3a* and *Dnmt3b*, were identified by Okano *et al.* (1999) and were proposed to be *de novo* methyltransferases, responsible for establishing DNA methylation patterns during embryonic development. They are highly expressed in ESCs but are downregulated upon differentiation, consistent with their proposed function in early development (Okano, Xie and Li, 1998; Okano *et al.*,

1999; Ramsahoye *et al.*, 2000). A degree of expression is retained in somatic cells however, with *Dnmt3a* expressed ubiquitously at low levels and *Dnmt3b* exhibiting higher expression levels in testes, thyroid and bone marrow (Xie *et al.*, 1999). Further differences between these highly homologous proteins were uncovered following their targeted disruption in mice. *Dnmt3a* deficient mice are viable but are runted and die within 4 weeks of birth (Table 1.1, Okano *et al.*, 1999). Global DNA methylation patterns appear to be unaffected in *Dnmt3a* null mice, however a progressive loss of germ cells in males is apparent, and a role for *Dnmt3a* in *de novo* methylation of imprints in germ cells was subsequently uncovered (Okano *et al.*, 1999; Kaneda *et al.*, 2004). The phenotype associated with *Dnmt3b* deficiency is more severe, as shown by the embryonic lethality of *Dnmt3b* null mice at around E14.5. Notable hypomethylation at minor satellites is apparent in *Dnmt3b* null mice, indicating defective *de novo* DNA methylation at these sequences in the absence of *Dnmt3b* (Okano *et al.*, 1999). The phenotypic differences of these mouse models demonstrate the important but distinct roles of *Dnmt3a* and *Dnmt3b* in embryonic development. Combined deletion of *Dnmt3a* and *Dnmt3b* is embryonic lethal at approximately E8.5, mimicking the *Dnmt1* null phenotype but displaying a lesser degree of hypomethylation (Okano *et al.*, 1999).

In addition to *Dnmt3a* and *Dnmt3b*, a new enzymatically active member of the *Dnmt3* family of *de novo* methyltransferases, *Dnmt3c*, was identified last year from a previously annotated pseudogene downstream of *Dnmt3b* (Barau *et al.*, 2016). *Dnmt3c* is exclusively expressed in male germ cells where it acts to methylate and silence evolutionarily young transposons, an activity that is required to preserve mouse fertility. This revelation highlights the complexity and plasticity of the *de novo* DNA methylation system and uncovers further mechanisms involved in retrotransposon silencing.

1.3.6.3 Maintenance vs de novo methylation

The classic view is that *de novo* methyltransferases *Dnmt3a* and *Dnmt3b* act to establish DNA methylation patterns in the early stages of development, which are then maintained to later stages of development by the maintenance methyltransferase *Dnmt1*. However, there is evidence to suggest that the functions of enzymes are not so

distinct. *Dnmt1* has been shown to have some *de novo* methyltransferase activity (Yoder *et al.*, 1997; Pradhan *et al.*, 1999). Futhermore, *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ double null ESCs which initially retain DNA methylation in early passages progressively lose almost all 5-mC during longer-term culture, suggesting an important role for *de novo* methyltransferases in maintenance of global DNA methylation, as well as establishment (Jackson *et al.*, 2004).

1.3.7 DNA methylation readers

As described previously, DNA methylation patterns are established and maintained by DNMTs. However, in mammals a specific set of proteins are employed to interpret these DNA methylation patterns, often referred to as DNA methylation 'readers'. These readers are known as methyl-CpG binding proteins (MBPs) of which there are three main families: the methyl-CpG binding domain (MBD) proteins (Lewis *et al.*, 1992; Meehan, Lewis and Bird, 1992; Hendrich and Bird, 1998), the Kaiso protein family (Prokhortchouk *et al.*, 2001; Filion *et al.*, 2006), and the SET and RING-finger associated (SRA) domain family (Unoki, Nishidate and Nakamura, 2004). The MBD family of proteins is the most well-characterised, with the identification of methyl-CpG binding protein 1 (MeCP1) first indicating the existence of methylation-specific binding proteins, followed by the cloning and characterisation of MeCP2 demonstrating their function in transcriptional repression *in vitro* (Meehan *et al.*, 1989; Lewis *et al.*, 1992; Meehan, Lewis and Bird, 1992). Many further members of the MBD protein family have since been identified, including MDB1-6 and SETDB proteins 1 and 2, which all contain a MBD domain after which they are named (Hendrich and Bird, 1998; Laget *et al.*, 2010). Although each MBD protein has a specific function, generally they act to repress transcription, at least in part through recruitment and targeting of multiple factors including HDACs, KMTs and chromatin remodelling factors, thus inducing heterochromatin formation and an inactive chromatin structure (Nan *et al.*, 1998; Ng, Jeppesen and Bird, 2000; reviewed by Du, 2015). MDB proteins have been shown to localise to pericentromeric heterochromatin, and are thought to contribute to constitutive heterochromatin formation and transcriptional silencing at these genomic regions (Hendrich and Bird, 1998; Ng, Jeppesen and Bird, 2000; Laget *et al.*, 2010).

1.3.8 DNA methylation co-factors

A number of non-methyltransferase proteins have been shown to be required to facilitate DNA methylation globally or locally, or in specific cellular contexts. These DNA methylation co-factors act to recruit, target or alter the chromatin environment to enable establishment or maintenance of DNA methylation. In this section I will discuss a few of the key DNA methylation co-factors that have been identified and characterised, including the putative chromatin remodelling helicase *Lsh* on which the majority of my PhD project is focussed.

1.3.8.1 Dnmt3l

A mammal-specific member of the *Dnmt3* family, *Dnmt3l* (Dnmt3-like) is a catalytically inactive co-factor to *de novo* DNA methylation. It has no known methyltransferase activity, but has been shown to bind to *Dnmt3a* and *Dnmt3b* to stimulate their methyltransferase activity *in vitro* (Gowher *et al.*, 2005). *Dnmt3l* is highly expressed in germ cells, where it is essential for *de novo* methylation and silencing of maternal and paternal imprinted loci and retrotransposons (Bourc'his *et al.*, 2001; Hata *et al.*, 2002; Bourc'his and Bestor, 2004; Kaneda *et al.*, 2004). Indeed, male mice lacking *Dnmt3l* are sterile due to defective spermatogenesis, possibly as a result of impaired *de novo* methylation and silencing of retrotransposons (Bourc'his and Bestor, 2004). *Dntm3l* deficiency in females does not cause sterility, but heterozygous offspring die mid-gestation and exhibit biallelic expression of imprinted genes due to lack of proper maternal imprint methylation (Bourc'his *et al.*, 2001). The phenotype of *Dnmt3l* null mice is strikingly similar to that of conditional *Dnmt3a* knockout in germ cells, indicating that both factors cooperate in the same mechanism to establish methylation at imprinted loci in germ cells (Kaneda *et al.*, 2004).

A more recent study from Neri *et al*. (2013) proposed a dual role for *Dnmt3l* in ESCs, where it is also highly expressed. Genome-wide analysis of DNA methylation in *Dnmt3l* knockdown ESCs revealed that *Dnmt3l* promotes DNA methylation at gene bodies while preventing methylation at the promoters of bivalent developmental genes. Disturbance of this methylation patterning by knocking down *Dnmt3l* prevented the differentiation of ESCs into primordial germ cells (PGCs), showing that *Dnmt3l* is required for this process (Neri *et al.*, 2013).

1.3.8.2 Uhrf1

The mechanism of maintenance methylation by *Dnmt1* is dependent on its E3 ubiquitin ligase co-factor *Uhrf1*, which is required to associate *Dnmt1* with replication foci and to maintain global and local DNA methylation levels following replication (Bostick *et al.*, 2007; Sharif *et al.*, 2007). *Uhrf1* contains a SRA domain which binds hemimethylated CpG dinucleotides, suggesting that *Uhrf1* acts to recruit *Dnmt1* to replication forks by binding hemimethylated CpGs during DNA replication (Sharif *et al.*, 2007; Arita *et al.*, 2008). More recently it has been demonstrated that binding H3K9me2/3 via the Tandem Tudor Domain (TTD) of *Uhrf1* can directly recruit *Dnmt1* to replication foci, linking H3K9 methylation and DNA methylation in the context of maintenance methylation (Liu *et al.*, 2013; Zhao *et al.*, 2016; Ferry *et al.*, 2017).

1.3.8.3 Lsh

Lsh (lymphoid specific helicase) is another protein that has been implicated as an important player in facilitating global DNA methylation. Murine *Lsh* was first cloned in T cell precursors by Jarvis *et al.* (1996), where it was shown to be highly expressed in B and T lymphocytes and associated tissues, which is why it was denoted the term *Lsh*. The human homologue of *Lsh* was cloned from human leukemic cells and was termed PASG (proliferation associated SNF2-like gene, Lee *et al.*, 2000), but *Lsh* is also known as HELLS (helicase, lymphoid specific) and SMARCA6 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 6). The majority of this thesis is concerned with the murine version of this gene, which will be referred to as *Lsh* and any references to the human homologue will use the term HELLS.

Lsh was identified as a member of the SNF2 subfamily of helicases, many of which are involved in chromatin remodelling activities (Geiman *et al.*, 2001). *Lsh* was shown to have a high level of similarity to another SNF2 subfamily member *DDM1* (decrease in DNA methylation) present in the flowering plant *Arabidopsis thaliana*. *DDM1* was previously identified in a forward genetics screen as a modulator of genomic DNA methylation in *Arabidopsis*, as *ddm1* mutants displayed a 70% loss in DNA methylation, particularly at repeated sequences (Jeddeloh, Stokes and Richards, 1999). Analysis of *lsh*$^{-/-}$ mice revealed a similar pattern of methylation loss, with *Lsh*-deficient

embryos displaying a 40-50% reduction in global DNA methylation levels, predominantly at satellite DNA and dispersed repetitive sequences (Dennis *et al.*, 2001). Throughout this section, I will discuss the function of *Lsh* as a co-factor to DNA methylation, focussing on the effects of *Lsh* removal on murine development, DNA methylation patterns and the transcriptional consequences. I will also discuss the potential epigenetic mechanisms of action for *Lsh* along with its contribution to development and disease.

1.3.8.3.1 Lsh gene structure and expression

The genomic structure and chromosomal localisation of *Lsh* was first described by Geiman *et al*. (1998), who showed it to be located on chromosome 19 in mice. *Lsh* is comprised of 21 coding exons and seven helicase domains and, based on the amino acid sequence of these helicase motifs, was identified as a member of the SNF2 subfamily of helicases (Figure 1.4). *Lsh* appears to be highly conserved in vertebrates, with homologues in yeast also apparent, as well as the *DDM1* homologue in *Arabidopsis*, of which there is 50% sequence identity with *Lsh* over the helicase domains (Geiman, Durum and Muegge, 1998; Meehan, Pennings and Stancheva, 2001). These seven helicase domains comprise multiple motifs critical for function such as the ATPase domain and DEAH (Asp-Glu-Ala-His) box (Eisen, Sweder and Hanawalt, 1995; reviewed by Clapier and Cairns, 2009).

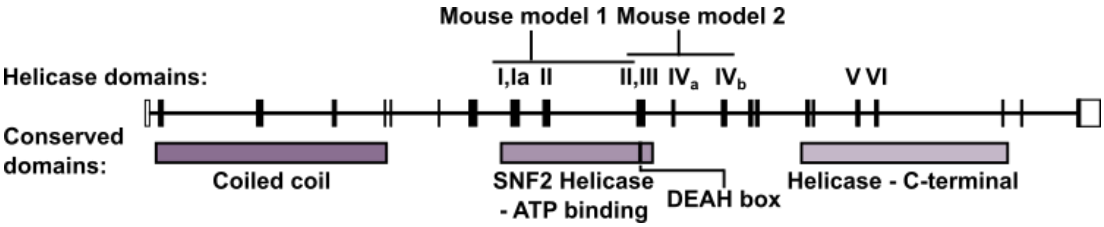


Figure 1.4. Schematic representation of murine *Lsh* gene. Coding exons are shown in black. Seven conserved helicase domains are indicated, as are conserved structural domains identified by UniProtKB. Helicase domains targeted by the two *Lsh* mouse models are also shown.

Lsh was initially identified as being expressed in highly proliferative B and T cells (Jarvis *et al.*, 1996). It was since shown to be more ubiquitously expressed across multiple mouse foetal tissues, suggesting a more general involvement of *Lsh* throughout the developing embryo outwith the lymphocyte lineages (Geiman *et al.*,

2001; Raabe *et al.*, 2001). In adult mice, *Lsh* is broadly expressed at low levels and enriched in tissues with high levels of proliferative and differentiating cells, such as the thymus, testis and bone marrow – an expression profile strikingly similar to that of *Dnmt3b* (Xie *et al.*, 1999; Raabe *et al.*, 2001). Furthermore, *Lsh* and HELLS have been shown to be highly expressed in ESCs, suggesting a role for *Lsh* in a pluripotent context (Assou *et al.*, 2007; Xi *et al.*, 2009).

1.3.8.3.2 Lsh mouse models

Two separate mouse models targeting *Lsh* for disruption have demonstrated that *Lsh* is essential for normal murine development and survival. In the first mouse model that was generated, helicase domains I, Ia and II were deleted *via* homologous recombination (mouse model 1 in Figure 1.4, Geiman *et al.*, 2001). These domains encompass the ATPase and DEAH box motifs, thought to be crucial for *Lsh* and SNF2 protein function (Figure 1.4). Western blotting confirmed the absence of any wild-type (WT) or truncated forms of LSH protein in $lsh^{-/-}$ mice. Absence of *Lsh* in developing mice resulted in perinatal lethality, with these mice dying a few hours after birth exhibiting reduced bodyweight (22% lower compared to WT littermates) and renal tubular defects. This demonstrated that *Lsh* was a non-redundant SNF2 family member whose function was critical for normal mouse development (Geiman *et al.*, 2001). Further analysis of this $lsh^{-/-}$ mouse model revealed specific defects in the generation of male and female germ cells. *Lsh* is essential for successful completion of meiosis in both oocytes and spermatocytes. De La Fuente *et al.* (2006) showed that while *Lsh* was not required for PGC formation and migration to the differentiating ovary, it was essential for the viability of primary oocytes in explanted ovaries. All mutant oocytes analysed showed incomplete chromosome synapsis, thus preventing recombination, establishment of crossovers and subsequent progression through meiosis. This study also revealed the dynamic localisation of *Lsh* throughout early meiosis, with transient localisation of *Lsh* to heterochromatic regions during synapsis followed by the association of *Lsh* along the length of the synapsed chromosomes in the pachytene stage of meiosis. These findings highlighted the crucial role *Lsh* plays in establishment of synapsis and progression of meiosis in female germ cells. This was further highlighted in a study demonstrating an essential role for *Lsh* in spermatogenesis, in

which *lsh*$^{-/-}$ male germ cells arrested in meiosis displaying chromosome synapsis abnormalities (Zeng *et al.*, 2011).

A second mouse model provided further evidence for an essential role for *Lsh* in postnatal development. Exons 10, 11 and 12 containing helicase domains III, IV and part of II were deleted, creating a hypomorphic mutation where a truncated version of LSH protein was expressed at low levels (mouse model 2 in Figure 1.4, Sun *et al.*, 2004). These mice display multiple physiological defects, resembling a premature ageing phenotype, and die prematurely within a few weeks of birth. Newborn mice exhibit a 25% reduction in bodyweight compared to littermates which is first apparent at E12.5 in the developing embryo. This reduced bodyweight becomes further exaggerated to a 70% difference in two-week-old mice. These mice display attributes of a premature ageing phenotype such as greying/loss of hair, reduced skin fat deposition, osteoporosis and kidney tubular degeneration. Approximately 60% die shortly after or within a few days of birth, while the remaining 40% survive several weeks after birth. Furthermore, fibroblasts derived from *Lsh* mutant embryos display a replicative senescence phenotype, with chromosomal abnormalities and karyotypic instability apparent in later passage fibroblasts. Characterisation of this mouse model provided further evidence for the vital role of *Lsh* in postnatal development, growth and longevity.

1.3.8.3.3 Lsh function in DNA methylation and transcription

1.3.8.3.3.1 Globally and at repetitive elements

Analysis of *Lsh* mutant mouse models revealed the fundamental role of *Lsh* in modulating genome-wide DNA methylation. Both models exhibited a substantial reduction in global DNA methylation levels, an approximate 50-60% decrease in *Lsh* null mice and a 30-40% reduction in hypomorphic *Lsh* mutant embryos, using methods based on the analysis of restriction enzyme sites (Dennis *et al.*, 2001; Sun, David W. Lee, *et al.*, 2004). High performance liquid chromatography (HPLC) analysis confirmed a 50% reduction in global cytosine methylation levels in *Lsh* null embryos (Tao *et al.*, 2011; Yu, Briones, *et al.*, 2014). DNA methylation levels were shown to be particularly reduced at satellite DNA and many dispersed repeat elements. Dennis *et al.* (2001) showed substantial hypomethylation at major and minor satellites as well

as IAPs, LINE-1, SINE-B1 and telomeric repeats in *Lsh* null embryonic tissues by Southern blotting. Analysis of *lsh*$^{-/-}$ oocytes also revealed extensive de-methylation of satellite DNA and IAP transposable elements, which was accompanied by reactivation of IAP transcription in *Lsh* deficient ovaries (De La Fuente *et al.*, 2006). Similar to the methylation patterns of *Lsh* null mice, considerable hypomethylation at minor satellites and IAPs accompanied by high levels of IAP expression was also apparent in hypomorphic *Lsh* mice (Sun, David W. Lee, *et al.*, 2004). Bisulfite sequencing in *lsh*$^{-/-}$ ESCs also revealed reduced DNA methylation at IAP elements, minor satellites and LINE-1 elements compared to WT, and that this hypomethylation was preserved during retinoic acid (RA)-induced differentiation (Ren *et al.*, 2015).

More recently, genome-wide sequencing approaches have defined DNA methylation defects at repetitive sequences in more detail. A report from Yu *et al.* (2014) showed a substantial decrease in DNA methylation across nearly all subclasses of repeats assayed by whole-genome bisulfite sequencing (WGBS) using WT and *lsh*$^{-/-}$ mouse embryonic fibroblasts (MEFs). A few instances of repeat hypermethylation in *lsh*$^{-/-}$ cells compared to WT were also observed. Transcriptional de-repression of a specific subset of the hypomethylated repeats including IAP elements was observed by RNA-seq, but the extent of hypomethylation did not correlate with the degree of reactivation of transcription. This indicates the presence of additional silencing mechanisms that act to repress hypomethylated repeats. A further study by Dunican *et al.* (2013) uncovered a more distinct signature of repeat hypomethylation in *Lsh*$^{-/-}$ fibroblasts using HELP-seq, a technique that generates high-resolution DNA methylation profiles informative of repeat regions. The largest impact of *Lsh* deletion on DNA methylation patterns was observed at LTR, LINE and satellite repeats, whereas methylation at other classes of repeats did not appear to be affected. RNA-seq expression analysis linked transcriptional reactivation of certain repeats, particularly IAP elements, to reduced DNA methylation levels in *lsh*$^{-/-}$ fibroblasts, while other hypomethylated repeats such as the LINE-1 elements remained largely transcriptionally repressed. This study also uncovered hypermethylation of some repeat classes, consistent with findings from Yu *et al.* (2014), suggesting that disruption of *Lsh* could act in part to redistribute genomic DNA methylation.

1.3.8.3.3.2 At single copy genes

As well as modulating DNA methylation at certain repeat elements, there is also evidence for *Lsh* facilitating methylation at selected genomic targets. Initial analysis of *lsh*$^{-/-}$ mice revealed a variable and context-dependent pattern of DNA methylation at low copy sequences. The imprinted *H19* gene was hypomethylated in *lsh*$^{-/-}$ mice compared to WT, whereas the imprinted *Igf2r* gene remained methylated. Additionally, the *β-globin*, *Pgk1* and *Pgk2* genes were hypomethylated in *lsh*$^{-/-}$ E13.5 embryos but not in newborn mice, indicating that the modulation of DNA methylation by *Lsh* may differ depending on developmental context (Dennis *et al.*, 2001). Fan *et al.* (2005) also reported hypomethylation of the promoter of the imprinted *Cdkn1c* gene in *lsh*$^{-/-}$ embryos, resulting in biallelic expression of this gene. Reduced DNA methylation was also apparent at the promoters of certain *Hox* loci in *lsh*$^{-/-}$ MEFs and correlated with transcriptional de-repression of these genes (Xi *et al.*, 2007; Tao *et al.*, 2010). A role for *Lsh* in regulation of DNA methylation at pluripotency genes has also been demonstrated. Xi *et al.* (2009) reported impaired establishment of DNA methylation at a subset of pluripotency genes including *Oct4* and *Nanog* during differentiation of *Lsh* knockdown embryonal carcinoma cells (ECCs). This resulted in incomplete silencing of these genes following differentiation, which was also observed in *lsh*$^{-/-}$ E8.5 embryos. However, this sustained *Oct4* expression was not evident in E18.5 *lsh*$^{-/-}$ embryos despite the continued absence of promoter methylation, suggesting the involvement of an alternative methylation-independent silencing mechanism at a later developmental stage. A further study identified the *Rhox* gene cluster as requiring *Lsh* for DNA methylation establishment at promoters and surrounding regions in *lsh*$^{-/-}$ MEFs, adding further weight to the idea that *Lsh* contributes to DNA methylation patterns during early developmental stages where *Rhox* genes are *de novo* methylated during lineage commitment (Myant *et al.*, 2011).

Genome-wide studies have provided further insight into the methylation status and distribution at unique sequences in *Lsh* mutant embryos and cells. The aforementioned study from Myant *et al.* (2011) utilised methyl-CpG binding affinity purification (MAP) combined with promoter-specific microarrays to demonstrate a reduction in DNA methylation of at least two-fold at approximately 20% of normally methylated

promoters in *lsh*$^{-/-}$ MEFs compared to WT. A small subset of promoters were also found to be hypermethylated in the absence of *Lsh*. This finding was reinforced by a study which produced a comprehensive map of DNA methylation distribution in *lsh*$^{-/-}$ E13.5 embryos using methylated DNA immunoprecipitation (MeDIP) combined with a whole-genome tiled array platform (Tao *et al.*, 2011). *Lsh* disruption resulted in changes to the DNA methylation profile at the 5' end of genes. In particular, a subset of single copy genes containing CpG islands at their 5' ends exhibited hypermethylation in *lsh*$^{-/-}$ embryos. This provided further evidence that *Lsh* can prevent as well as promote DNA methylation, and that *Lsh* acts to regulate the deposition and distribution of methylation at unique sequences. Lastly, WGBS of *lsh*$^{-/-}$ MEFs provided high-resolution cytosine methylation maps which revealed a greater loss in DNA methylation at low CpG density promoter regions of unique protein-coding and noncoding RNA genes (Yu, McIntosh, *et al.*, 2014).

1.3.8.3.3.3 Over large chromosomal domains

In addition to informing about DNA methylation distribution across genes, the genome-wide studies discussed above all provided novel insights into the modulation of DNA methylation by *Lsh* over large chromosomal regions. Tao *et al.* (2011) reported large regions with well-defined boundaries that were differentially methylated (hypo- and hyper-methylated) in *lsh*$^{-/-}$ embryos over chromosomes 8 and 15. This suggests that distinct chromosomal domains are regulated by *Lsh* in tandem using a common epigenetic mechanism. Myant *et al.* (2011) also noted clusters of promoters on all autosomes whose DNA methylation patterns appear to be regulated by *Lsh*. Moreover, high-resolution WGBS maps generated by Yu *et al.* (2014) demonstrated the extent and diversity of genomic loci differentially methylated in *lsh*$^{-/-}$ MEFs compared to WT, suggesting that *Lsh* targets a nuclear compartment rather than specific genomic sequences. These differentially regulated regions consistently overlapped with lamin B1 attachment domains (LADs), indicating that *Lsh* function may involve targeting this nuclear compartment for DNA methylation regulation.

Collectively, these studies highlight complex nature of DNA methylation modulation by *Lsh*. The major targets of *Lsh* are repetitive elements, particularly satellite DNA and LTRs. However, *Lsh* also participates in regulation of DNA methylation patterns

at unique sequences, especially those involved in pluripotency, development and lineage-commitment. Furthermore, there is increasing evidence that *Lsh* targets specific nuclear compartments, chromosomal domains and clusters of genes to regulate epigenetic patterns over large chromosomal regions rather than at discrete genomic loci. The extent and diversity of the sequences affected by *Lsh* depletion hints at a sophisticated mechanism for *Lsh* function which may include cooperation with DNMTs, chromatin and other epigenetic modulators. These potential mechanisms will be discussed in a subsequent section.

1.3.8.3.4 Potential mechanisms of Lsh function

It is clear that *Lsh* modulates genome-wide DNA methylation in mice, but the molecular mechanisms underlying this process remain unclear. It is still not fully understood whether *Lsh* acts directly or indirectly on genomic loci to regulate DNA methylation and whether any chromatin remodelling activities are involved. Furthermore, it is still disputed as to whether *Lsh* acts as a recruitment factor for DNMTs or associates with other epigenetic factors as part of a complex or scaffold to modify DNA methylation. The potential mechanisms of *Lsh* function will be overviewed in this section.

1.3.8.3.4.1 Lsh as a co-factor for DNMTs

The amino acid sequence of *Lsh* harbours no recognisable methyltransferase domain, suggesting that *Lsh* does not act to directly methylate genomic targets. Rather it is hypothesised that *Lsh* recruits or directs the activity of DNMTs to establish or maintain DNA methylation patterns at selected genomic loci (Meehan, Pennings and Stancheva, 2001; Dunican *et al.*, 2013). Disruption of *Lsh* in mice resulted in remarkable global hypomethylation in the absence of any apparent changes in DNMT expression levels or activity (Dennis *et al.*, 2001). However, *Lsh* has been shown to interact and cooperate with *Dnmt1*, *Dnmt3a* and *Dnmt3b* in *vitro* (Zhu *et al.*, 2006; Myant and Stancheva, 2008; Xi *et al.*, 2009; Dunican, Pennings and Meehan, 2015). Much evidence supports the current working hypothesis that *Lsh* predominantly contributes to *de novo* DNA methylation. *Lsh* was shown to be required for DNA methylation establishment on episomal vectors introduced into MEFs, while it was not required for methylation maintenance of already methylated episomes (Zhu *et al.*, 2006). This

study also demonstrated that *Lsh* affects both *Dnmt3a* and *Dnmt3b*-mediated *de novo* methylation and silencing of a retroviral transgene, and confirmed a direct interaction between *Lsh* and both *de novo* DNMTs by co-immunoprecipitation. Further studies have implied that *Lsh* primarily acts in concert with *Dnmt3b*, as *Lsh* depletion results in reduced association of *Dnmt3b* with specific genomic targets such as *Hox* and stem cell gene promoters (Xi *et al.*, 2007, 2009). Two recent reports have highlighted the importance of *Dnmt3b* for *Lsh* function. Termanis *et al.* (2016) demonstrated *de novo* methylation of repeats and gene promoters in $lsh^{-/-}$ MEFs when *Lsh* was re-introduced and that this methylation re-establishment was impaired upon *Dnmt3b* knockdown. Ren *et al.* (2015) further highlighted the cooperation of *Dnmt3b* and *Lsh,* showing that *Dnmt3b* binding to and *de novo* methylation of specific repeat elements (such as IAPs) is dependent on the presence of *Lsh*. These reports indicate that *Lsh* and *Dnmt3b* act in concert to target specific genomic regions for *de novo* DNA methylation. However, comparison of $lsh^{-/-}$ and $dnmt3b^{-/-}$ methylomes illustrated that *Lsh* can also function independently of *Dnmt3b* to facilitate DNA methylation deposition and transcriptional silencing of selected repeats, such as LTRs (Dunican *et al.*, 2013).

Largely, however, it appears that the probable role of *Lsh* in *de novo* DNA methylation primarily involves *Dnmt3b*, although the functional requirements for this role remain unclear. The crucial function of the ATP binding site and DEAH motifs became apparent in the aforementioned studies form Termanis *et al.* (2016) and Ren *et al.* (2015). These reports demonstrated that *de novo* methylation of repetitive elements and unique sequences in ESCs and MEFs relied on fully functional LSH protein with an intact ATP binding site and DEAH motif. Furthermore, the targeting of *Dnmt3b* to repeat loci was shown to be dependent on the presence of WT *Lsh,* as targeting was impaired in ATP and DEAH mutant LSH protein (Ren *et al.*, 2015). These observations highlight a critical role for the ATPase activity of *Lsh* in directing *Dnmt3b* to establish cytosine methylation patterns at selected repeats and single copy sequences, hinting towards the involvement of *Lsh* in chromatin remodelling activities.

1.3.8.3.4.2 Lsh as a chromatin remodeller

SNF2 family members are characterised by their ability to disrupt histone-DNA interactions. This can act to alter chromatin structure, in part through ATP-dependent

nucleosome sliding, influencing the accessibility of enzymes to DNA. *Lsh* is a putative chromatin remodeller based on its amino acid homology to other SNF2 family members, yet chromatin remodelling activity by *Lsh* has yet to be demonstrated. However *DDM1*, the homologue of *Lsh* in *Arabidopsis*, has been shown to induce nucleosome repositioning on a short DNA fragment *in vitro* in an ATP-dependent manner (Brzeski and Jerzmanowski, 2003). Due to the homology of *Lsh* to the SNF2 family, it has been proposed that *Lsh* may also exert nucleosome remodelling activity to promote heterochromatin structure while regulating access of *de novo* methyltransferases to certain genomic loci (Meehan, Pennings and Stancheva, 2001). Recent WGBS analysis of *DDM1* mutants has reinforced this theory. DNA methylation was markedly depleted at linker histone 1 (H1) rich DNA in *ddm1* mutants, demonstrating a specific requirement for *DDM1* for methylation at regions of densely packed chromatin (Zemach *et al.*, 2013). It has been demonstrated *in vitro* that DNA wrapped around nucleosomes is protected from DNA methylation, and that the presence of SNF2 proteins improves *de novo* methylation *in vitro* by remodelling nucleosomes to provide a better substrate for *de novo* DNMTs (Felle *et al.*, 2011). Therefore it has been proposed that *DDM1*, and perhaps *Lsh*, promotes access of DNMTs to nucleosomal DNA at H1-rich heterochromatin, incurring DNA methylation deposition at these regions (Zemach *et al.*, 2013).

Heterochromatin is the major target of *Lsh*. *Lsh* is known to be required for DNA methylation mainly at lowly expressed genes and repetitive sequences which are embedded in heterochromatin (Yu, McIntosh, *et al.*, 2014). It has also been referred to as an 'epigenetic guardian' of repetitive elements due to the multiple epigenetic mechanisms *Lsh* appears to employ to maintain silent chromatin at these repeats (Huang *et al.*, 2004). *Lsh* has been shown to strongly associate with heterochromatin, particularly at pericentromeric regions, in an ATP-dependent manner (Yan *et al.*, 2003; Lungu *et al.*, 2015). This suggests that the putative chromatin remodelling function of *Lsh* is required for its activity at heterochromatin. A recent study aimed to investigate the nucleosome remodelling capabilities of *Lsh* using the Nucleosome Occupancy and Methylome (NOMe) sequencing assay, which provides a high resolution nucleosome footprint alongside a DNA methylation profile. This report suggested that *Lsh* is required to maintain nucleosome density at repeat sequences, as nucleosome

occupancy was reduced at IAP and LINE-1 elements in *lsh*$^{-/-}$ ESCs compared to WT following RA-induced differentiation (Ren *et al.*, 2015). WT levels of nucleosome occupancy were restored upon re-introduction of fully functional LSH protein, but not ATPase and DEAH mutants. This led the authors to propose that nucleosome remodelling is the principal function of *Lsh*, acting to influence patterns of DNA methylation at repetitive elements. Another very recent study further elucidated a potential mechanism for *Lsh* function in chromatin remodelling using a tethering strategy to target a GAL4-LSH fusion protein to an engineered *Oct4* locus in ESCs. Prior to differentiation, *Lsh* associates with the *Oct4* locus, reducing the active H3K4me3 histone mark and chromatin accessibility without any effects on DNA methylation or gene expression. Upon differentiation, *Lsh* association induces transcriptional repression along with an increase in repressive H3K9me3 and DNA methylation (Ren *et al.*, 2017). Along with the insights into the remodelling capacity of *DDM1*, this sheds light on an exciting and potentially crucial function of *Lsh* at heterochromatin. Indeed, the report from Ren *et al.* (2017) illustrates a potential mechanism for *Lsh* in priming specific genes in ESCs by changing chromatin accessibility. This results in their eventual repression during differentiation due to *Lsh* promoting heterochromatin formation, and subsequent reinforcement of this repression through the acquisition of DNA methylation.

1.3.8.3.4.3 Lsh association with complexes and histone modifications

There is also evidence that *Lsh* may interact and cooperate with protein complexes to influence histone modifications and chromatin accessibility. *Lsh* was shown to associate with PRC1 components in MEFs and was required for proper establishment of PRC1-mediated H2A-K116 patterns at *Hox* genes, with effects on PRC2-mediated H3K27me3 also apparent (Xi *et al.*, 2007). More recent genome-wide reports have also noted aberrant patterns of H3K27me3, as well as other marks such as H3K4me3 and H3K9me3 in *lsh*$^{-/-}$ MEFs (Dunican *et al.*, 2013; Yu, McIntosh, *et al.*, 2014). *Lsh* is thought to cooperate with H3K9me3 to reinforce heterochromatin formation and transcriptional silencing at repeats, and an intact H3K9me3 pathway is required for release of *Lsh* from pericentromeric heterochromatin (Dunican *et al.*, 2013; Yu, McIntosh, *et al.*, 2014; Lungu *et al.*, 2015). *Lsh* has also been proposed to cooperate

with the G9a/GLP complex responsible for most H3K9 dimethylation at chromatin. The absence of *Lsh* resulted in impaired recruitment of G9a/GLP to *Rhox* promoters, hampering the establishment of silenced chromatin and transcriptional repression at these loci (Myant *et al.*, 2011). Increased H3 acetylation was also evident at these *Rhox* genes, leading to the hypothesis that *Lsh* may bind and promote deacetylation, providing a suitable substrate for H3K9 dimethylation by G9a/GLP. This proposed connection of *Lsh* with H3 deacetylation is backed up by previous studies demonstrating cooperation of *Lsh* with HDACs: *Lsh* recruits HDACs and DNMTs to aid the formation of transcriptionally repressive chromatin and *Lsh* dissociates from pericentromeric heterochromatin following prolonged treatment with HDAC inhibitors (Yan *et al.*, 2003; Myant and Stancheva, 2008). Taken together, these reports indicate that *Lsh* functions in concert with an array of proteins and complexes to alter histone modifications, chromosome accessibility, gene expression and DNA methylation.

1.3.8.3.5 Lsh in pluripotency and development

Much of the research that has been undertaken into *Lsh* function has pointed to a role for *Lsh* in establishment of DNA methylation at genes involved in development and pluripotency (Xi *et al.*, 2007, 2009; Myant *et al.*, 2011). There have also been insights into the functional requirement for *Lsh* in preserving proper pathways of lineage commitment and differentiation in mice. *Lsh* has been shown to be necessary for normal haematopoesis, spermatogenesis and development of oocytes (De La Fuente *et al.*, 2006; Fan *et al.*, 2008; Zeng *et al.*, 2011). Furthermore, the involvement of *Lsh* in pluripotency and differentiation has been investigated, demonstrating the *Lsh*-dependent *de novo* methylation and repression of a subset of stem cell-specific genes and the maintenance of stem-cell like characteristics in *Lsh*-depleted ECCs following differentiation (Xi *et al.*, 2009).

The involvement of *Lsh* in regulation of stem cell gene expression has been further demonstrated in two recent reports from Ren *et al.* (2015 and 2017) that were discussed previously. These revealed an important potential function for *Lsh* in inducing DNA methylation and heterochromatin formation to silence certain genes, such as the pluripotency factor *Oct4*, during ESC differentiation. These indications that *Lsh* plays

a role in influencing pluripotency factor expression, coupled with the high expression of *Lsh* in ESCs compared to somatic cells, makes stem cell systems an attractive model in which to study the contribution of *Lsh* to DNA methylation during development.

Recently, there have been two in-depth studies into the differentiation capacity of *lsh*$^{-/-}$ ESCs, further elucidating the role of *Lsh* in pluripotency and lineage commitment. Yu *et al.* (2014) observed an appearance of H3K4me1, a mark of putative enhancer elements, clustered at neuronal lineage genes that were hypomethylated in *lsh*$^{-/-}$ MEFs compared to WT. These differentially enriched H3K4me1 regions were maintained upon reprogramming to iPSCs, where the neuronal genes they marked showed increased expression. This resulted in the enhanced propensity of *lsh*$^{-/-}$ iPSCs to differentiate towards the neural lineage pathway, exemplifying the link between *Lsh*, *de novo* methylation and cellular plasticity *via* perturbation of H3K4me1 patterns. A further study has recently described an apparently contrary effect of *Lsh* disruption on neural differentiation. Han *et al.* (2017) showed that the absence of *Lsh* led to defects in the differentiation of neural stem progenitor cells (NSPCs), such as delayed lineage commitment and reduced neural lineage marker expression in mature cultures. This study also demonstrated a critical role for *Lsh* in regulating the proliferation and self-renewal of NSPCs *via* its association with enhancers of key stem cell and cell cycle regulators *Bmp4* and *Cdkn1a*. Reports that *Lsh* interacts with core pluripotency factors such as *Oct4* adds weight to the idea that *Lsh* could play a role in regulating self-renewal through interaction with the pluripotency network (van den Berg *et al.*, 2010). However, further studies are required to uncover how this regulation may influence lineage commitment and differentiation, with particular focus on the role of *Lsh* in priming and progression of neural differentiation.

1.3.8.3.6 Lsh in cancer and disease

Lsh mutant mouse models have demonstrated a crucial role for *Lsh* in normal murine development and survival. Recently, the human homologue of *Lsh*, HELLS, was found to be mutated in the rare autosomal recessive disease ICF syndrome (Thijssen *et al.*, 2015). This disorder is largely regarded as an immunodeficiency disorder, as the life expectancy of ICF patients is significantly reduced due to recurrent, severe respiratory and gastrointestinal infections that are often fatal. This is accompanied by a failure to

thrive, developmental delay, distinctive facial abnormalities and varying degrees of intellectual disability (Hagleitner *et al.*, 2007; Weemaes *et al.*, 2013). Pericentromeric chromosomal instability is a hallmark of this disorder, manifesting in characteristic rearrangements of chromosomes 1, 9 and 16, regions which exhibit substantial DNA methylation loss and chromosome decondensation in ICF patients (Jeanpierre *et al.*, 1993; Hagleitner *et al.*, 2007). This reinforces the indications from *Lsh*-deficient mouse models that *Lsh* is required for DNA methylation and heterochromatin maintenance at pericentromeric regions, and demonstrates the physiological consequences of loss of this function in humans. Missense mutations in *Dnmt3b* account for approximately 50% of ICF cases, providing further support to the proposal that *Lsh* predominantly functions alongside *Dnmt3b* to facilitate DNA methylation deposition, particularly at repetitive regions of the genome (Hansen *et al.*, 1999; Xu *et al.*, 1999; Ehrlich *et al.*, 2008). Interestingly, four out of the five mutations in HELLS identified in ICF patients were outwith the conserved helicase domains, indicating crucial functions for less well-characterised regions of the protein (Thijssen *et al.*, 2015).

Aberrant HELLS expression and its epigenetic function has also been associated with development and progression of cancer. Increased HELLS expression has been shown to be characteristic of a diverse range of cancers and is implicated in playing a critical role in tumour growth and metastasis, ultimately driving cancer progression (Keyes *et al.*, 2011; von Eyss *et al.*, 2012; Wang *et al.*, 2015; He *et al.*, 2016). HELLS exerts this action through association with other epigenetic factors, such as G9a and the lncRNA HOTAIR, affecting the expression of key genes and tumour suppressors (Wang *et al.*, 2015; He *et al.*, 2016). The widespread hypomethylation caused by *Lsh* deletion in mice has also been shown to promote erythroleukemia development and is associated with genomic instability (Fan *et al.*, 2003, 2008). Another recent report has provided further insight into the contribution of *Lsh* to genome stability. Litwin *et al.* (2017) found that the *Saccharomyces cerevisiae* homologue of *Lsh*, Irc5, interacts with the cohesin complex, and that disruption of *IRC5* led to reduced accumulation of cohesin on chromosomes. This resulted in premature sister chromatid separation, which has implications for genomic stability, as well as for accurate DNA replication and chromosome segregation during mitosis. These studies highlight physiological roles

for *Lsh*, as well as the critical function in epigenetic regulation of cell growth and genomic stability.

1.4 Epigenetic mechanisms in embryonic development

Mammalian embryonic development is the process by which a single-cell zygote undergoes cell division and differentiation, resulting in a complex and multi-cellular mammalian embryo. Development of an embryo from a zygote is a complex process requiring co-ordination of multiple global and cell-specific mechanisms to direct changes in gene expression, while the underlying DNA sequence remains unchanged. These gene expression changes are primarily orchestrated by a network of core transcription factors (TFs), *Oct4*, *Sox2* and *Nanog* (reviewed by Li and Belmonte, 2017). Epigenetic mechanisms act alongside TF networks to regulate TF binding and to reinforce expression changes, directing cell fate decisions and promoting developmental progression.

Establishment of stable, heritable epigenetic marks directs lineage commitment, progressively restricting the developmental potential of cells and establishing barriers to prevent reversion to preceding cellular states. However, genome-wide epigenetic patterns can be extensively reprogrammed to reinstate developmental potency at two key stages during mammalian development - in the zygote after fertilisation and in PGCs (Smith *et al.*, 2012; reviewed by Hackett and Surani, 2013). Extensive global epigenetic remodelling can also be induced *in vitro*, exemplified by generation of iPSCs from somatic cells (Takahashi and Yamanaka, 2006). These systems highlight the importance of dynamic epigenetic changes during development in defining cell state and developmental plasticity.

Multiple epigenetic systems including DNA methylation, histone modifications and small RNAs co-operate to influence the progression of embryonic development. DNA methylation undergoes possibly the most extensive changes during development, and its importance in development is highlighted by the relatively early embryonic lethality of the DNMTs (Li, Bestor and Jaenisch, 1992; Lei *et al.*, 1996; Okano *et al.*, 1999). Specifically, 5-mC has been demonstrated to be critical for directing and maintaining cell differentiation in embryonic lineages, as hypomethylated *Dnmt1*$^{-/-}$ ESCs are viable

and proliferate normally, but undergo p53-mediated apoptosis when induced to differentiate (Li, Bestor and Jaenisch, 1992). Due to the essential role of DNA methylation in influencing and maintaining cellular identity, the mechanisms which erase and establish DNA methylation during development and the direct effects of these epigenetic events are of particular interest and are still not fully understood.

1.4.1 Epigenetic dynamics during early embryonic development

During the early stages of mammalian embryonic development, DNA methylation undergoes dramatic global reprogramming. This ensues in the zygote upon fertilisation, where there is rapid and extensive de-methylation of the genome (illustrated in Figure 1.5). It has been suggested that the zygotic DNA de-methylation is TET enzyme mediated (primarily TET3) and proceeds through oxidation of 5-mC to 5-hmC, followed by a passive loss of 5-hmC through DNA replication (reviewed by Hill, Amouroux and Hajkova, 2014). However, detailed analysis indicates that 5-mC is removed from the paternal genome prior to 5-hmC accumulation and that inactivation of TET3 does not interfere with this process; implying that the processes underlying DNA de-methylation is complex and still to be fully explained (Amouroux *et al.*, 2016; reviewed by Nashun, Hill and Hajkova, 2015). This progressive de-methylation culminates in functionally pluripotent cells of the inner cell mass (ICM) of the blastocyst (approximately E3.5 in mice) that display a globally hypomethylated genome. It is this stage of development from which ESCs are derived, which will be discussed in more detail in section 1.4.3. Some genomic regions retain DNA methylation in the hypomethylated ICM, such as IAPs and genomic imprints, however some dispersed repeats such as LINEs and selected LTR classes undergo dramatic de-methylation (Lane *et al.*, 2003; Kim *et al.*, 2004).

Following genome-wide DNA methylation erasure, global DNA methylation patterns are progressively re-established in the ICM in the implanting blastocyst, during the morphogenetic transformation of the blastocyst into a cup-shaped epithelium known as the egg cylinder (Figure 1.5). This wave of DNA methylation deposition assists in directing lineage specification and ultimately maintaining cellular identity and genomic stability. *De novo* DNA methyltransferases *Dnmt3a* and *Dnmt3b* and their co-factor *Dnmt3l* direct this global re-methylation of the genome in the implanting

blastocyst. These patterns of DNA methylation are then maintained throughout all subsequent cell divisions by *Dnmt1* and its co-factors. Genome-wide studies using reduced representation bisulfite sequencing (RRBS) and meDIP indicated that the vast majority of *de novo* DNA methylation re-establishment is completed prior to the onset of gastrulation at E6.5 (Borgel *et al.*, 2010; Smith *et al.*, 2012). A study examining *de novo* DNA methylation during trophoblast differentiation highlighted gains of DNA methylation across specific lineage-specific promoters such as *Elf5* that act as an early epigenetic restriction and a determinant of cell fate (Ng *et al.*, 2008). A further study also demonstrated the establishment of 5-mC at specific pluripotency and germ-line specific gene promoters during mESC progression to lineage-committed neural progenitors (Mohn *et al.*, 2008). This acquisition of DNA methylation early in the differentiation process correlated with lineage commitment, suggesting a role for *de novo* DNA methylation in repression of pluripotency and preventing differentiation down the germ cell lineage. Accumulation of DNA methylation at key promoters exemplifies the idea that progressive DNA methylation acts to influence early developmental decisions and ultimately restrict cell fate, reminiscent of Waddington's depiction of the epigenetic landscape that influences and restricts developmental potential.

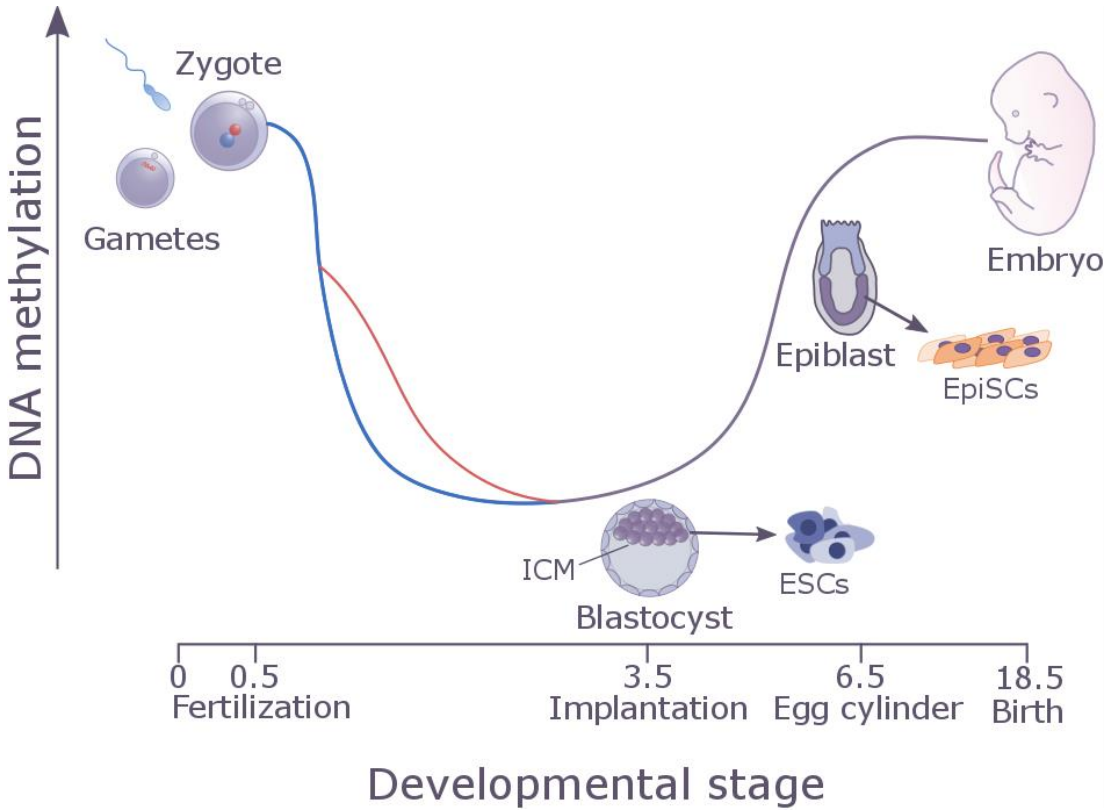


Figure 1.5. DNA methylation remodelling during embryonic development. Rapid DNA de-methylation ensues after fertilisation of the zygote in the paternal genome (represented in blue), with DNA de-methylation progressing at a steadier pace in the maternal genome (shown in red). This genome-wide erasure of 5-mC culminates in globally hypomethylated cells of the ICM in the blastocyst, from which ESCs are derived. Genome-wide patterns of DNA methylation are re-established during implantation of the blastocyst. The majority of 5-mC patterns are re-established by day E6.5, resulting in a hypermethylated genome in the post-implantation epiblast (the stage from which EpiSCs are derived).

A second wave of global de-methylation occurs during PGC specification between E8.5 and E12.5. The genome-wide hypomethylation observed in PGCs is more extensive than that of pre-implantation ICM, as it includes erasure of genomic imprints so that they can be re-established in a sex-specific manner during PGC specialization (Hajkova *et al.*, 2002). In contrast to hypomethylated ICM cells, DNA methylation is lost at almost all repetitive elements except IAPs were it is partially retained (Lane *et al.*, 2003). Re-methylation ensues from E12.5 during differentiation and maturation of gametes, although the dynamics of *de novo* DNA methylation differ between male and female germ cells. This results in re-establishment of distinct DNA methylomes between the hypermethylated sperm and partially methylated oocyte genomes (Smith

et al., 2012). These mature gametes may ultimately go on to fuse, restarting the mammalian life cycle from the zygotic stage.

1.4.2 Pluripotent stem cells as a model for early development

Pluripotent stem cells are defined as having the potential to generate almost every cell type in the body. They possess the ability to generate any of the three primary germ layers: ectoderm, mesoderm and endoderm, and can therefore give rise to almost all the tissues and organs that make up an adult organism. Pluripotent stem cells do however retain a degree of lineage restriction, as they do not efficiently contribute to the trophoectoderm or primitive endoderm, unlike totipotent cells which can generate all extra-embryonic tissues (Beddington and Robertson, 1989).

There are several key types of pluripotent stem cells, with the most commonly used forms being ESCs derived from the ICM of the blastocyst, embryonic germ cells (EGCs) isolated from the primordial germ layer of a developing embryo, and ECCs that can be established from teratomas that arise in the ICM. Isolation and culture of these pluripotent stem cell models has provided valuable tools in which to study development *in vitro*. More recently, a breakthrough discovery showed that adult somatic cells can be directly (genetically) reprogrammed into iPSCs through ectopic expression of specific pluripotency factors *Oct4*, *Sox2*, *Klf4* and c-*Myc* (Takahashi and Yamanaka, 2006). This discovery provided a great deal of promise for the areas of developmental biology, regenerative medicine and epigenetics. For this thesis, I focussed on the most commonly used and best characterised pluripotent stem cell type, ESCs, as an *in vitro* model for early mammalian embryonic development.

1.4.3 Isolation and culture of mouse ESCs

ESCs were originally isolated and cultured from the ICM of the mouse blastocyst in 1981 by two groups (Evans and Kaufman, 1981; Martin, 1981). These mouse ESCs (mESCs) were shown to have the capacity to self-renew, meaning that they could be established as progressively growing cultures *in vitro.* The pluripotent status of these mESCs was confirmed when they were shown to be able to differentiate into a wide variety of cell types that contribute to all three germ layers. Despite the transiency of

the pluripotent state that cells of the ICM occupy *in vivo*, their *in vitro* mESC counterparts can be maintained indefinitely under the appropriate culture conditions. Even after extensive culture, mESCs retain key attributes of ICM cell identity and potency and can contribute to all adult tissues in chimeric animals following reintroduction into the developing blastocyst (Bradley *et al.*, 1984). The capability of mESCs to re-enter the developmental program in chimeric animals to produce functional somatic and germ cells was pivotal for the development of techniques to produce transgenic mouse models. Combination of these ESC techniques with homologous recombination technologies paved the way to targeted transgenesis, enabling precise manipulation of the mouse genome, and so the production of knock-out, knock-in and conditional transgenic mouse and cell lines (Doetschman *et al.*, 1987; Thomas and Capecchi, 1987; Downing, Battey Jr. and Battey, 2004). The key features of self-renewal and developmental potency also make mESCs a valuable tool for modelling development and disease mechanisms *in vitro*.

Conventionally, mESC cultures are maintained on a layer of mitotically inactivated 'feeder' fibroblasts and in the presence of foetal calf serum, which provide previously undefined factors enabling the sustained proliferation of these cultures whilst maintaining their pluripotent status. The cytokine leukaemia inhibitory factor (LIF) was later identified as the crucial factor secreted by fibroblasts, a discovery which enabled the generation of mESC cultures maintained in LIF and serum in the absence of feeders (Smith *et al.*, 1988; Williams *et al.*, 1988). LIF promotes pluripotency by inducing expression of pluripotency genes, such as *Klf4*, through binding to the gp130 receptor to activate transcription factor STAT3 in a JAK kinase-dependent manner (Niwa *et al.*, 1998, 2009). Alongside LIF, bone morphogenic protein 4 (BMP4), which is present in foetal calf serum, was identified as the other key component that supports mESC self-renewal and supresses neural differentiation in culture (Ying *et al.*, 2003; Zhang *et al.*, 2010). It achieves this by acting via SMAD signalling pathways to stimulate expression of *Inhibition of differentiation* (*Id*) genes (Ying *et al.*, 2003). Together, LIF and BMP4 (provided by serum) constitute the central components required to protect the pluripotent status of mESC cultures through both induction of pluripotency networks and suppression of differentiation.

The ‘conventional culture’ of mESCs in LIF and serum as described above can lead to some issues with the reproducibility of cultures due to the presence of undefined and variable factors within and between batches of animal-derived serum. The mechanisms by which LIF and serum act to maintain mESCs can also affect the stability of cultures. The mode of action of LIF and serum focuses on counteracting the downstream effects of differentiation priming. Along with activation of undefined signalling pathways by unregulated serum components, this creates an environment comprising numerous and often conflicting signals (reviewed by Hackett and Surani, 2014). Although this does not affect the functional pluripotency of the culture as a whole, it results in a population of cells occupying a spectrum of pluripotent states. This manifests in cultures displaying heterogeneity in terms of morphology, gene expression and differentiation capacity (Chambers *et al.*, 2007; Hayashi *et al.*, 2008; Toyooka *et al.*, 2008). Some cells will resemble a state of ‘naïve’ or ‘ground state’ pluripotency, representative of cells in the ICM of the blastocyst, while others may more closely reflect a ‘primed’ state, associated with expression of lineage specification markers and are proposed to be more developmentally advanced (Hayashi *et al.*, 2008; Canham *et al.*, 2010). mESCs cultured in serum and LIF are dynamic, with cells continually cycling between these pluripotent phases (Toyooka *et al.*, 2008; Abranches *et al.*, 2013). Thus, these mESC cultures can be described as ‘metastable’. This metastability has been proposed by some to be a functional property of mESC cultures, allowing preparation for lineage commitment whilst preserving the pluripotent status of the population as a whole (Torres-Padilla *et al.*, 2014). Others argue that the heterogeneity of mESCs is a result of sub-optimal conventional culture conditions and that there is a lack of evidence for the existence of transcriptional heterogeneity in the epiblast *in vivo* (Kalkan and Smith, 2014).

Additionally, and crucially, populations of mESCs cultured in the presence of serum and LIF exhibit global DNA hypermethylation, despite being isolated from the hypomethylated ICM (Meissner *et al.*, 2008; Habibi *et al.*, 2013; Leitch *et al.*, 2013). The average CpG methylation of cultured mESCs in serum is approximately 80%, a level similar to those of primed or lineage-restricted cells (Stadler *et al.*, 2011). Indeed, WGBS indicated that DNA methylation profiles of male mESCs resemble those of post-implantation E6.5 epiblast (Habibi *et al.*, 2013). This highlights that the DNA

methylomes of cultured mESCs may not faithfully reflect those of the pre-implantation ICM from which they are derived, and so may not represent the ideal model for investigation of DNA methylation patterns in early development.

1.4.4 The ground state of pluripotency

To overcome the issues associated with conventional mESC derivation and culture, more well-defined serum-free culture systems were sought. In 2008, Ying *et al.* described the use of two small molecule kinase inhibitors ('2i') which simultaneously block *Fgf/Erk* signalling and enhance *Wnt* signalling, resulting in a population of cells occupying a state of naïve pluriopotency. Specifically, small molecule PD0325901 (PD03) targets *Mek* for inhibition, blocking the differentiation-inducing consequences of *Erk* signalling, whereas CHIR99021 (CHIRON) promotes pluripotency through *Gsk3* inhibition, which stabilises β-catenin, mimicking canonical *Wnt* signalling. Collectively, the 2i components stabilise and promote naïve pluripotency by insulating mESCs from differentiation signalling and inducing pluripotency and self-renewal networks (Ying *et al.*, 2008). This results in a more uniform expression of key pluripotency genes such as *Nanog*, *Esrrb* and *Prdm14* as well as reduced expression of lineage-associated factors, resulting in a population of mESCs that are transcriptionally and functionally distinct from those cultured in serum and LIF (Figure 1.6, Marks *et al.*, 2012).

Crucially, establishment of naïve pluripotency also coincides with global hypomethylation. Analysis of global 5-mC levels using mass spectrometry techniques revealed an approximate three-fold reduction in global DNA methylation in mESCs cultured in 2i and LIF compared to their serum counterparts (Habibi *et al.*, 2013; Leitch *et al.*, 2013). Genome-wide methylome analysis showed that the distribution of DNA methylation in 2i/LIF mESCs was comparable to the hypomethylated cells of the ICM from which ESCs are derived, indicating that the 5-mC profiles of 2i mESCs may more faithfully reflect methylation patterns present in the pre-implantation epiblast (Ficz *et al.*, 2013; Habibi *et al.*, 2013; Choi *et al.*, 2017). Some genomic loci were resistant to the de-methylation associated with transition to naïve pluripotency, such as IAP elements and genomic imprints, consistent with the unchanged expression levels of *Dnmt1* which are required to maintain methylation at these regions (Ficz *et al.*, 2013;

Habibi *et al.*, 2013; Leitch *et al.*, 2013). The retention of 5-mC at these regions parallels those which escape the wave of de-methylation *in vivo* following fertilisation of the zygote. Somewhat surprisingly, when 2i/LIF mESCs were returned to serum/LIF culture, global 5-mC levels fully recovered and showed appropriate re-establishment of DNA methylation patterns (Habibi *et al.*, 2013; Leitch *et al.*, 2013). This illustrates the interconvertibility between these culture states and the distinct DNA methylomes that they confer, highlighting the remarkable plasticity of the supposedly relatively stable 5-mC mark in this culture system (illustrated in Figure 1.6). The interconvertibility of this system also provides opportunities to dynamically modify DNA methylation levels in mESC culture models.

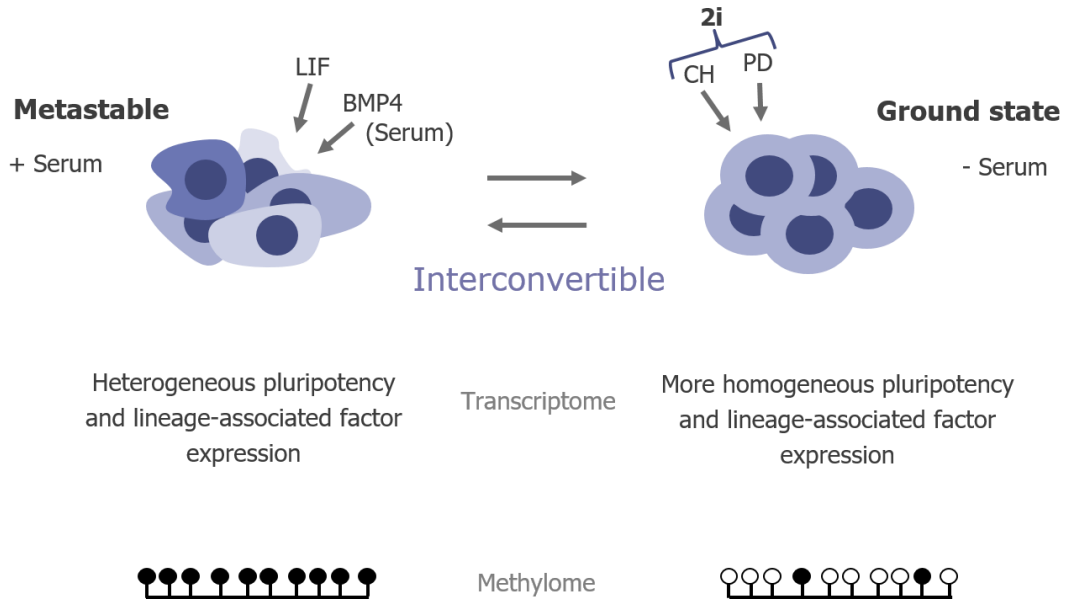


Figure 1.6. Schematic depicting the characteristic features of serum and 2i cultures and the interconvertibility between these pluripotent states. 'Metastable' mESCs cultured in the presence of serum exhibit heterogeneous expression of pluripotency and lineage-associated factors and are globally hypermethylated. They can be reversibly transitioned to the pluripotent 'ground state', where pluripotency factor expression becomes more uniform, lineage-associated factor expression is reduced, and the genome exhibits global hypomethylation.

A caveat of using the 2i culture system as a model for the ICM is that perturbation of *Wnt* and *Fgf/Erk* signalling is linked to genome instability and cellular transformation in ESCs (Acevedo *et al.*, 2007; Raggioli *et al.*, 2014; Chen *et al.*, 2015). An increase in karyotypic abnormalities has been reported following prolonged culture (more than 20 passages) of mESCs in 2i (Hassani *et al.*, 2014). A recent study has also

demonstrated irreversible epigenetic perturbations, such as progressive de-methylation of ICRs at most imprinted genes, alongside chromosomal aberrations after long-term 2i culture, affecting the developmental potential of these mESCs (Choi *et al.*, 2017). Therefore, maintenance of mESCs in the presence of 2i is not advised, and short-term (less than 10 passages) culture of mESCs is preferred for modelling DNA methylation in pre-implantation development.

The mechanisms governing the global DNA methylation reprogramming observed during transition from primed serum mESCs to naïve pluripotency have remained a topic of discussion since its discovery. Ficz *et al.* (2013) proposed that de-methylation occurs through a combination of active and passive mechanisms. Within the first 72 hours of 2i transition, initial global 5-mC erasure was accompanied by a two-fold transient enrichment in 5-hmC, suggesting that active removal of 5-mC through a 5-hmC intermediate partly accounts for DNA de-methylation in the early stages of 2i transition. This initial active DNA de-methylation was followed by replicative loss of 5-mC and 5-hmC throughout the genome during the remainder of the transition to ground state pluripotency. This passive de-methylation was thought to be assisted by the disabling of *de novo* DNA methylation through rapid transcriptional repression of *Dnmt3a* and *Dnmt3b* and their co-factor *Dnmt3l*, while expression levels of *Dnmt1* and its co-factor *Uhrf1* remained constant (Ficz *et al.*, 2013). Other reports supported this model of cooperation between active and passive mechanisms alongside *de novo* methylation suppression, resulting in genome-wide 5-mC erasure associated with transition to 2i culture (Habibi *et al.*, 2013; Leitch *et al.*, 2013). Indeed, a critical role for PR-domain-containing 14 (*Prdm14*) in 2i-associated de-methylation was revealed by Yamaji *et al.* (2013), who demonstrated that *Prdm14* downregulates the expression of *de novo* DNMTs via recruitment of the repressive complex PRC2, and that the presence of *Prdm14* is essential for global hypomethylation during the transition to the pluripotent ground state. The reversion of ground state mESCs into serum/LIF culture results in the upregulation of *Dnmt3a*, *Dnmt3b* and *Dnmt3l* expression, which is thought to drive the re-establishment of DNA methylation throughout the genome (Habibi *et al.*, 2013; Leitch *et al.*, 2013; Yamaji *et al.*, 2013).

A recent report employed mathematical modelling of DNA de-methylation kinetics to dissect the contribution of different de-methylation mechanisms to 2i-associated genome-wide hypomethylation. This analysis implicated impaired DNA methylation maintenance in global 5-mC erasure, rather than active de-methylation or disrupted *de novo* methylation. A reduction in *Dnmt1*-targeting co-factor *Uhrf1* alongside a genome-wide loss of H3K9me2, which is thought to be required for association of *Uhrf1* with chromatin, is proposed to disrupt targeting of *Dnmt1* to replicating DNA, resulting in impaired 5-mC maintenance and its subsequent reduction throughout the genome of ground state mESCs (von Meyenn *et al.*, 2016). These conflicting models for 5-mC erasure during transition to naïve pluripotency highlight the further investigation that is required to uncover the mechanistic regulation of this DNA de-methylation pathway.

Alongside the extensive remodelling of DNA methylation that occurs upon transition towards naïve pluripotency, there is a genome-wide reconfiguration in the patterns of repressive histone modifications. The histone mark deposited by PRC2, H3K27me3, is globally redistributed in ground state mESCs. Unlike in conventional serum mESCs where H3K27me3 is enriched over the promoters of PRC target genes, H3K27me3 is depleted at promoters in 2i, although this does not coincide with transcriptional activation of the associated genes. Total cellular levels H3K27me3 are comparable between serum and 2i mESCs, suggesting a redistribution of the mark rather than global depletion, which is confirmed by a substantial increase of H3K27me3 apparent at major satellite regions (Marks *et al.*, 2012). A report from Walter *et al.* (2016) that utilised 2i in combination with vitamin C to deplete 5-mC even more extensively than in 2i/LIF media alone also observed an accumulation of H3K27me3 at repetitive elements, suggesting that this repressive mark is sequestered to these regions to reinforce silencing of transposons in the absence of DNA methylation. Furthermore, the redistribution of H3K27me3 away from promoters results in a marked reduction of bivalency in ground state mESCs, implying that fewer genes are poised for development in this pluripotent state (Marks *et al.*, 2012). Along with reconfiguration of H3K27me3 deposition, there is also evidence of other epigenetic perturbations, such as a global reduction in H3K9me3, and particularly H3K9me2, in the pluripotent ground state (Leitch *et al.*, 2013; von Meyenn *et al.*, 2016; Walter *et al.*, 2016).

1.4.5 Culture models representing later developmental stages

The cluster of 10-20 cells that are present in the ICM of the blastocyst give rise to more than 200 specialised cell types that are present in the adult organism. The biological processes of lineage commitment and differentiation that underlie the generation of this diverse array of tissues from a relatively small pool of pluripotent cells is an area of active research. Over the years, methods involving *in vitro* differentiation of mESCs have been developed and optimised. This has allowed dissection of the molecular and epigenetic mechanisms involved in the biological processes of lineage commitment and differentiation, as well as providing culture models representing the early stages of post-implantation development.

A key and widely-used differentiation method involves LIF withdrawal and aggregation of mESCs in suspension, resulting in the formation of structures called embryoid bodies (EBs). The differentiation of mESCs to EBs can generate all three germ layers, as well as wide variety of specialised cell types such as those in the cardiovascular, hepatic and neural lineages (Doetschman *et al.*, 1985; Burkert U, von Rüden T, 1991; Murry and Keller, 2008). Addition of growth factors to mESC cultures during transition to EBs can also influence the route of differentiation, with addition of specified concentrations of RA promoting formation of neurons (Wichterle *et al.*, 2002). The process of EB formation is thought to reflect the orderly dynamics of *in vivo* differentiation, as EBs have been reported to exhibit a degree of self-organisation and appear to undergo a gastrulation-like process during their generation (ten Berge *et al.*, 2008). They have been demonstrated to form a primitive streak-like region, where cells undergo an epithelial-to-mesenchymal transition and differentiate into mesodermal progenitors, and appear to establish an anteroposterior polarity. This renders EBs an effective model in which to study the gastrulation process *in vitro*, and has added advantages over differentiation protocols involving adherence to matrices or in monolayers due to the ability of EBs to form gastrulation-like structures and polarity.

More recently, cell lines were derived from post-implantation mouse epiblasts by two groups using culture conditions including *Fgf* and *Activin* in the absence of LIF, conditions also sufficient for long-term propagation of human ESCs (Brons *et al.*,

2007; Tesar *et al.*, 2007). These cultures, termed epiblast stem cells (EpiSCs), express the core pluripotency factors *Oct4*, *Sox2* and *Nanog*, but all other naïve pluripotency genes are transcriptionally silent (Guo *et al.*, 2009). Instead, markers of the early post-implantation epiblast are expressed such as *Fgf5*, along with lineage-specific genes such as endodermal *Foxa2* and mesodermal *Brachyury*. However, EpiSC populations are highly heterogeneous, as is the expression of lineage-specific factors, with EpiSC populations proposed to occupy a range of states representing the early epiblast to up to the late-gastrula-stage embryo (Han *et al.*, 2010; Kojima *et al.*, 2014; Tsakiridis *et al.*, 2014). EpiSCs have been proposed to be true *in vitro* counterparts of the post-implantation epiblast as they retain functional properties of the epiblast, such as the ability to differentiate into somatic lineages *in vitro* and can generate chimeras when grafted into post-implantation embryos in culture (Han *et al.*, 2010; Huang *et al.*, 2012; Kojima *et al.*, 2014).

EpiSCs can also be generated in culture from ESCs propagated in the presence of *Fgf2* and *Activin* (Guo *et al.*, 2009). This involves multiple passages to stabilise the culture after a period of substantial cell death. However, parallels can be drawn between this ESC to EpiSC conversion and the *in vivo* differentiation process as one copy of the X-chromosome becomes epigenetically silenced in female cells, suggesting that similar epigenetic mechanisms are active in both systems. Additionally, EpiSCs generated in culture represent a developmentally restricted state as they do not spontaneously revert to the pluripotent state without genetic manipulation (Guo *et al.*, 2009). Although these EpiSCs can be inefficiently reprogrammed to naïve pluripotency with the introduction of a single pluripotency factor *Klf4*, they are proposed to represent a distinct developmental, epigenetic and functional state from naïve ESCs in culture, and provide an *in vitro* model of the pre-gastrulation epiblast (Guo *et al.*, 2009; Nichols and Smith, 2009).

1.5 Thesis aims

The erasure and re-establishment of DNA methylation during early embryonic development in order to acquire then progressively restrict cellular potential remains a remarkable example of the power of epigenetics. However, despite the great interest in these biological processes, the mechanisms involved in governing these transitions

are still not fully understood. It is becoming increasingly clear that the complement of factors involved in enacting epigenetic processes in a cell is dependent on cellular identity and developmental context. Therefore, I aimed to dissect the contribution of selected factors to the epigenetic mechanism of DNA re-methylation during the developmental transitions observed in the early stages of cellular potential restriction (i.e. differentiation).

The DNA methylation co-factor and putative chromatin remodelling helicase *Lsh* has been conclusively shown to be essential for normal levels of DNA methylation in mouse embryos. Its crucial role in development had also been demonstrated, as *Lsh*-depleted mice exhibit multiple physiological defects and die either perinatally or in their infancy. However, most previous studies of *Lsh* function had focussed on the effects of *Lsh* depletion in somatic cells and embryonic tissues, with little insight into the requirement for or function of *Lsh* in the early stages of development, where DNA methylation patterns are set in accordance with lineage commitment. I set out to decipher the contribution of *Lsh* to DNA methylation in these developmental stages using culture models where establishment of DNA methylation and restriction of developmental potential could be assessed. Therefore, the main aims of my thesis were specified as follows:

- To develop and optimise a mESC system in which to dynamically modulate DNA methylation levels, allowing observation of DNA methylation establishment in a pluripotent context
- To determine the requirement for *Lsh* function in DNA methylation establishment in this mESC culture system
- To investigate the contribution of *Lsh* to DNA methylation deposition in culture models representing later developmental stages, during the restriction of cell fate potential associated with differentiation

Chapter 2. Materials and Methods

2.1 Mammalian cell culture

All mammalian cells were cultured on cell culture grade tissue culture flasks (Corning) or plates (nunclon delta surface, Thermo Scientific) in a humidified atmosphere at 37°C and 5% CO2.

2.1.1 Cell lines and media

E14 and CRISPR *lsh*$^{-/-}$ mESCs generated from E14 mESCs were used throughout this project (CRISPR *lsh*$^{-/-}$ line generated by Dr Donncha Dunican, unpublished). WT and *dnmt3l*$^{-/-}$ mESCs generated using TALEN gene editing were kindly provided by the Oliviero group (Neri *et al*., 2013). WT and HELLS mutant HAP1 cell lines were purchased from Horizon Genomics (www.horizondiscovery.com). The media recipes for all cell lines used are detailed in Tables 2.1 and 2.2.

	Serum	**EB**	**SNLP**	**HAP1**	**Company**
Media type	GMEM	DMEM	DMEM	IMDM	Gibco
FCS	15%	20%	15%	10%	In-house
NEAA	1%	1%	1%	-	Sigma
Sodium pyruvate	1%	-	1%	-	Sigma
L-glutamine	1%	1%	1%	1%	In-house
Pen/strep	1%	1%	1%	1%	In-house
β-mercaptoethanol	50 µM	100 µM	-	-	Invitrogen
LIF	500 units/ml	-	-	-	Millipore

Table 2.1. Culture media recipes. Companies that reagents were acquired from are indicated. In-house reagents were prepared by HGU technical services.

	2i	EpiSC	Company
Basal Media			
Neurobasal	50%	50%	Invitrogen
DMEM-F12	50%	50%	Invitrogen
N2-supplement	0.5x	0.5x	Invitrogen
B27-supplement	0.5x	1x	Invitrogen
7.5% BSA	0.05%	-	Invitrogen
Pen/strep	1%	1%	In-house
NEAA	-	1%	Sigma
β-mercaptoethanol	-	50 µM	Invitrogen
Complete Media			
LIF	1000 units/ml	-	Millipore
L-glutamine	1%	1%	In-house
PD0325901	1 µM	-	Stemgent
CHIR99021	3 µM	-	Stemgent
Monothioglycerol	0.15 mM	-	Sigma
FGF2	-	10 ng/ml	R&D systems
Activin A	-	20 ng/ml	Peprotech

Table 2.2. Specialised culture media recipes. Basal media recipe is indicated in the top portion of the table. Complete media refers to the components that were added to 50 ml aliquots of the basal media and used within 3 days of preparation.

2.1.2 mESC serum and 2i culture

The 'serum' and '2i' media that mESCs were cultured in are detailed in Tables 2.1 and 2.2. LIF was included in both serum and 2i media. Serum and 2i mESCs were passaged every 2-4 days upon reaching ~80-90% confluency. To passage, the media was removed and the cells rinsed with 1x PBS (Sigma). The cells were detached from the culture surface using 1x trypsin-EDTA v/v (Sigma) and gentle agitation following incubation at 37°C for 2-3 minutes. The trypsin was then inactivated by adding media containing serum then centrifuging at 1000 rpm for 5 minutes. For 2i cells, an additional wash step was performed to remove traces of serum using GMEM only medium (Gibco). Cells were then resuspended in the appropriate medium and seeded into new flasks or plates coated with 0.2% gelatin (G1890, Sigma, prepared in 1x PBS). Media was changed when required, or every 1-2 days for mESCs cultured in 2i.

2.1.3 Transition to epiblast stem cells (EpiSCs)

E14 and *lsh*$^{-/-}$ mESCs converted to 2i in the presence of LIF for at least 15 days (treated as passage 0) were transitioned to EpiSCs in the absence of LIF using the media detailed in Table 2.2. Culture flasks/plate surfaces were prepared by adding 0.75 µg/cm^2 of fibronectin diluted in 1x PBS (Millipore, FC010). This was incubated at room temperature for 10 minutes, then removed and surfaces were left to dry in a sterile culture hood for at least 30 minutes prior to use. At the start of the transition, 2 x 10^5 2i mESCs were seeded per well of a 6-well plate coated with fibronectin in 2i medium. The cells were left to attach to the plate for approximately 6 hours before the medium was changed to 'EpiSC' medium. During the transition, cells were passaged upon reaching ~80-90% confluency, which was between every 2-7 days depending on the stage of the transition. To dissociate the cells, 1x accutase (StemPro) was added to the cells after rinsing with 1x PBS. The cells were carefully collected in EpiSC medium and centrifuged at 1000 rpm for 5 minutes, then gently resuspended in EpiSC medium and seeded into new plates or flasks coated with fibronectin. Media was changed every 1-2 days during the transition to EpiSCs.

2.1.4 Differentiation to embryoid bodies (EBs)

E14 and *lsh*$^{-/-}$ mESCs adapted to 2i in the presence of LIF for 15 days were transitioned to EBs using the media detailed in Table 2.1. On day 0 of the protocol, 20 µl droplets of 'EB' media (without LIF) containing 200 or 400 2i mESCs were plated onto the lids of sterile 10 cm^2 dishes (Thermo Scientific) using a multichannel pipette. On each lid, 48 droplets were plated. The plates were filled with 20 ml 1x PBS to maintain a humid environment for the hanging drop EBs to form in. The plates were left undisturbed until day 3, when the newly-formed EBs were removed from the lids and placed into non-gelatinised 10 cm^2 sterile dishes to achieve further growth in suspension. EB medium was used to flush the EBs from two plate lids (approximately 96 EBs) into a 50 ml Falcon tube. The EBs were then left to settle to the bottom of the tube at room temperature for 15 minutes. Excess media was aspirated and the EBs were resuspended in fresh EB medium and placed into a non-gelatinised 10 cm^2 dish. Care was taken not to disaggregate the EBs during this process, helped by using stripettes with a large opening (25 ml or 50 ml). The 200- and 400-cell EBs were

cultured separately. On day 5, the 200- and 400-cell EBs were collected and excess media removed using the same procedure as on day 3. The 400-cell EBs from one dish were resuspended in fresh EB medium and seeded into a 10 cm^2 culture dish (nunclon delta surface, Thermo Scientific) coated with 0.2% gelatin. The 200-cell EBs were seeded into gelatinised 12-well culture plates containing one glass coverslip per well, aiming to place 4-5 EBs per well. At this stage, RA (Sigma) was added to the media of the appropriate wells/plates at a final concentration of 1 µm. The EBs settled and attached to these culture surfaces and were incubated a further 7 days. The media was carefully changed every 2-3 days. On day 12, the EBs in the 10 cm^2 dishes were harvested (as described in section 2.1.4.) using trypsin-EDTA to detach them from the culture surface. The EBs cultured on coverslips were fixed *in situ* (as described in Section 2.6.1.) for analysis by immunofluorescence.

2.1.5 Blastocyst derivation and culture

Blastocysts were isolated from *Lsh* conditional knockout mice with the aim of deriving *lsh*$^{-/-}$ mESC lines to confirm results obtained using CRISPR-generated *lsh*$^{-/-}$ lines. Two mice conditional for the genetrap construct in the *Lsh* locus were crossed and E3.5 blastocysts isolated from the resulting pregnant female. Blastocysts were collected, plated and disaggregated (at day 7) as described in Czechanski *et al*. (2014) in 2i media (containing LIF, shown in Table 2.2) on a layer of mitotically inactivated SNLP feeders. Established mESC lines were maintained in 2i with LIF and on feeders and were passaged using the method described in section 2.1.1., using a seeding density of 1-3 x 10^5 cells per cm^2. Transgenic mouse lines were maintained and crosses set up by David Read. Blastocyst collection, plating and disaggregation was carried out by Dr Ian Adams.

2.1.5.1 Preparation of mitotically inactivated feeders

A layer of mitotically inactivated SNLP feeders were required to support derivation and growth of mESC lines derived from blastocysts. SNLPs were cultured in the media described in Table 2.1. They were treated with 10 µg/ml mitomycin C (Sigma) for 2.5 hours, then washed thoroughly twice with PBS before dissociation with 1x trypsin.

These mitotically inactivated SNLPs were then counted and seeded at a density of 1 x 10^4 cells per cm^2 for use in blastocyst derivation and culture.

2.1.6 Recovery and cryopreservation of mammalian cell lines

To recover a cell line, the appropriate vial was retrieved from liquid nitrogen and stored in dry ice until thawing in a 37°C water bath. The cell suspension was carefully added to 10 ml of the appropriate culture media, then centrifuged at 1000 rpm for 5 minutes. The resulting cell pellet was resuspended in the required culture media before seeding into culture dishes/flasks with the appropriate coating.

To maintain stocks of cell lines, cells that were surplus to requirements during passaging were prepared for cryopreservation in liquid nitrogen. Following dissociation and centrifugation, cell pellets were resuspended in CryoStor cell cryopreservation media CS10 (Sigma) to obtain cell concentrations of 2-5 x 10^6 cells per ml and added to cryovials. These were subsequently stored at -80°C overnight in a Mr Frosty (ThermoFisher Scientific) to facilitate a steady rate of freezing (-1 °C a minute) for optimal cell preservation. Following overnight storage at -80 °C, the vials were transferred to liquid nitrogen (-140°C) for long-term storage.

2.1.7 Counting and harvesting of mammalian cell lines

For cell counting, cultures were dissociated following the appropriate method, centrifuged at 1000 rpm for 5 minutes and the resulting pellet thoroughly resuspended in 2-5 ml of the appropriate medium to achieve a single-cell suspension, 10 µl of which was pipetted into a haemocytometer counting chamber. The number of cells in all four quadrants were counted, divided by 4 to give the average number of cells per quadrant, then multiplied by 1 x 10^4 to give the number of cells per ml of suspension. This allowed me to seed the correct number of cells at the beginning of the EpiSC and EB transition experiments.

To prepare cell cultures for harvesting, cells were dissociated using the appropriate method, collected in media then centrifuged at 1000 rpm for 5 minutes. The resulting pellet was resuspended in 1x PBS then centrifuged again at 1000 rpm for 5 minutes. All supernatant was removed and pellets for DNA or protein were snap frozen on dry

ice and stored at -80°C. Pellets for RNA were resuspended in 1 ml Trizol reagent (Invitrogen), snap frozen on dry ice and stored at -80°C.

2.2 DNA preparation, manipulation and analysis

2.2.1 DNA extraction from mammalian cell pellets

Cell pellets were resuspended in 1 ml lysis buffer (100 mM Tris pH 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl) supplemented with 2 µl RNaseA cocktail (Ambion) and incubated for approximately 1 hour at 37 ºC. Following this incubation, 5 µl of 20 mg/ml Proteinase K (ThermoFisher Scientific) was added and the samples were incubated at 55ºC overnight with shaking at 1000 rpm in a thermomixer.

The next day an equal volume of UltraPure Phenol:Chloroform:Isoamyl Alcohol (PCI, 25:24:1 v/v, Invitrogen) was added, the sample was mixed vigorously then centrifuged at 4000 rpm for 15 minutes at 4 ºC. The upper aqueous phase containing the DNA was pipetted off from the lower organic phase containing protein and cell debris. This was added to a new tube and a second PCI extraction was performed. A final chloroform only (Sigma) extraction was performed and the aqueous phase collected in a new eppendorf. One fifth of the sample was transferred to a new eppendorf, 2.5 volumes of ice-cold 100% ethanol were added along with 1/10th volume 3 M sodium acetate (Sigma). The extractions were then incubated at -20 ºC for at least 1 hour to overnight to precipitate the DNA. The precipitate was then centrifuged at 13,000 rpm at 4 ºC for 5 minutes to pellet the DNA. The supernatant was discarded and the pellet washed with 70% ethanol and centrifuged again at 13,000 rpm at 4 ºC for 5 minutes. The pelleted DNA was left to air dry for approximately 5 minutes before resuspension in 30-50 µl mass spectrometry grade water (LC-MS Chromasolv, Sigma). These samples were stored at -20 ºC for downstream analysis by liquid chromatography mass spectrometry to determine methylcytosine content. To the remainder of the PCI extracted sample, 2-3 volumes of ice-cold 100% ethanol were added and the sample gently inverted to mix. A microbiological loop was used to spool out the precipitated DNA. The DNA was washed with 70% ethanol and air dried for 3-5 minutes. The DNA was resuspended in 200-300 ul Tris-EDTA (TE, 100 mM Tris pH 8, 10 mM EDTA pH 8) and stored at 4 ºC or -20 ºC.

2.2.2 Nucleic acid quantification

The concentration was measured and purity assessed of both dsDNA and ssRNA spectrophotometrically using a Nanodrop ND-1000. 1.5 µl of sample was loaded and the A_{260} absorbance measured to calculate the concentration of the sample (as an OD of 1 at 260 nm = 50 µg/ml of dsDNA). The $A_{260/280}$ and $A_{260/230}$ ratios were also assessed to determine the purity of the nucleic acids.

2.2.3 Agarose gel electrophoresis

Nucleic acids were resolved using UltraPure agarose (Invitrogen) gels made using 1x TBE (Tris/Borate/EDTA) or 1x TAE (Tris/Acetic acid/EDTA) and supplemented with 1x SYBR safe (ThermoFisher Scientific) or 1 µg/ml ethidium bromide (EtBr, Invitrogen). DNA samples were loaded using orange G loading buffer (50% glycerol v/v; 5 mM EDTA pH 8; 0.3% orange G w/v). GeneRuler DNA ladder mix (ThermoFisher Scientific) was used to determine the size of DNA fragments. Agarose gels were visualised using a UV transilluminator or a FLA-5100 imaging system (FujiFilm Life Science).

2.2.4 Restriction digestion of genomic DNA

Restriction digestion using methylation-sensitive enzymes was used to assess the extent of CpG methylation at their restriction sites, which are enriched in repetitive sequences in the genome. Methylation-sensitive restriction endonucleases *MaeII (HpyCHIV,* NEB), *HpaII* (NEB) were used along with a methylation-insensitive isoschizomer of *HpaII*, *MspI* (NEB). 1 µg of RNA-free genomic DNA was incubated with 10 units of each of these enzymes and 1x CutSmart buffer (NEB) in a total volume of 20 µl at 37 ºC for 2 hours. Digested DNA was resolved on a 1.5% agarose-TBE gel run at 100-150 volts for 1.5-2 hours. The gel was then stained in 1 µg/ml EtBr and imaged using the FLA-5100 imaging system.

2.2.5 Southern blotting

Southern blotting in combination with restriction digestion was used to estimate DNA methylation levels at major and minor satellite genomic regions. Genomic DNA samples restriction digested with *MaeII*, *HpaII* or *MspI* were run on a large 1% agarose-TBE gel at 100 volts for approximately 3 hours (for assessment of major satellites) or 50 volts for approximately 6 hours (for minor satellites). The gel was post-stained with 1 µg/ml EtBr and imaged using the FLA-5100 imaging system. The gel was washed 2 x 15 minutes in denaturing solution (0.5M NaOH, 1.5M NaCl) and then 2 x 15 minutes in neutralising solution (0.5M Tris pH7.5, 1.5M NaCl) on a rocking platform at room temperature. The gel was equilibrated in 20x saline-sodium citrate (SSC) for at least 10 minutes before the DNA was transferred onto a Hybond-N+ membrane (GE Healthcare) overnight by capillary action. The membrane and gel were placed between 1 mm Whatman paper saturated with 20x SSC, with the ends of the paper dipped in a reservoir of 20x SSC. Dry paper towels and a weight were placed on top of the upper Whatman to draw the buffer through the stack, transferring the DNA onto the membrane. Following transfer, DNA was crosslinked to the membrane using the Stratalinker 1500 UV crosslinker (Stratagene). Hybridisation and probe labelling was performed using AlkPhos Direct Labelling and Detection System (Amersham). Membranes were pre-hybridised at 55 ^{o}C with rotation in 20 ml of the provided hybridisation buffer supplemented with 0.8 g of the provided blocking reagent (4% w/v) and 2 ml 5M sodium chloride (0.5M final concentration). 50 ng of purified probe DNA (detailed in section 2.2.5.1) was denatured in a boiling waterbath for 5 minutes, placed on ice for 5 minutes then labelled as follows:

5 µl 10 ng/µl denatured probe DNA
5 µl Reaction buffer
1 µl Labelling reagent
5 µl Cross-Linker solution (diluted 1 in 4 in water)

The reaction was incubated at 37 ^{o}C for 30 minutes then added to the membrane in the hybridisation buffer and incubated overnight at 55 ^{o}C with rotation (final probe concentration was 2.5 ng/ml). The probed membrane was washed 2 x 10 minutes at 55 ^{o}C with primary wash buffer (2M Urea, 0.1% w/v SDS, 50 mM Na phosphate pH 7, 150 mM NaCl, 1 mM $MgCl_2$, 0.2% w/v Blocking reagent) then washed 2 x 5 minutes

at room temperature in 1x secondary wash buffer (20x = 1M Tris base, 2M NaCl, pH 10) supplemented with $MgCl_2$ to a final concentration of 2 mM. For detection, 7 ml CDP-Star detection reagent (Amersham) was added to the membrane for 5 minutes, then developed and imaged using an ImageQuant LAS 4000 (GE Healthcare).

2.2.5.1 Probe preparation

Plasmid DNA for major satellite (pSAT, Lewis *et al.*, 1992) and minor satellite (R198, Kipling *et al.*, 1994) genomic regions was double restriction digested (*EcoRI/BamHI*, NEB, for major satellite, *EcoRI/HindIII*, NEB, for minor satellite) to excise the fragment to use as a Southern blot probe. 2 µg plasmid DNA was digested at 37 ºC for 2 hours with 20 units of each of the appropriate enzymes in a total reaction volume of 50 µl with 1x CutSmart buffer (NEB). The fragment for the probe (major satellite 240 bp, minor satellite 360 bp) and the plasmid DNA were separated on a 1% agarose-TAE gel and the fragment was gel purified as described in Section 2.5.2. The plasmids used to prepare the probes for Southern blotting were kindly provided by Professor Nick Gilbert.

2.2.6 Mass spectrometry analysis of 5-methylcytosine content

Liquid chromatography mass spectrometry (LC-MS) was used to identify and quantify global levels of 5-methylcytosine (5-mC) in DNA samples from mammalian cells. Using this technique, genomic DNA was hydrolysed to single nucleotides which were then extracted using methanol and acetonitrile and separated on a spectrometry column, using the monoisotropic mass in daltons (Da) to identify desired nucleotide. Standards for cytosine and 5-mC (Zymo) were prepared in the same way as and run alongside the samples for comparison. For results shown in Chapter 3 and Appendix Figure A2 the following protocol was used for digestion and LC-MS analysis of genomic DNA:

Degradation of DNA to single nucleotides:

1 µg genomic DNA (in mass spectrometry grade water)
2.5 µl DNA degradase buffer (1x, Zymo research)
1 µl DNA degradase enzyme (10 units, Zymo research)
Up to 25 µl with mass spectrometry grade water

The reaction was incubated at 37 ºC for 2 hours followed by 70 ºC for 20 minutes to heat-inactivate the enzyme. It was then centrifuged at 10,000 rpm for 20 minutes to spin down the majority of the enzyme to the bottom of the tube. The upper 10 µl of the solution was removed and added to 40 µl of a 5:3 mixture of methanol and acetonitrile (mass spectrometry grade, Sigma) to give a 5:3:2 mixture of methanol:acetonitrile:sample. This was centrifuged at 10,000 rpm for a further 5 minutes at room temperature. The upper 40 µl was removed and submitted to the IGMM mass spectrometry facility for LC-MS analysis performed by Jimi Wills as follows: 10 µl of the provided sample was separated on a SeQuant ZIC-pHILIC (150x4.6mm) column using a Thermo UltiMate 3000 BioRS, with a flowrate of 0.3 mL/min and a gradient from 90% to 5% B in 10 minutes, where B is acetonitrile and A is 20 mM ammonium carbonate. Mass spectra were acquired in negative mode on a Thermo Q Exactive, scanning from 300 to 350 m/z at resolution 70k. AGC target was set to 3 x 10^6 and maximum ion time 500 ms. LC-MS data were analysed using Excalibur software to assess the quality of the mass spectra peaks and to quantify the area under the peaks.

For results shown in Chapter 4, LC-MS analysis methods had been improved and so the following protocol was used: 2.5 µg genomic DNA was made up to 44 µl in mass spectrometry grade water and incubated at 95 ºC for 10 minutes. 5 µl 10x T7 DNA polymerase reaction buffer and 1 µl 10 units/µl T7 DNA polymerase (Thermo Fisher Scientific) were added and the mixture incubated overnight at 37 ºC. The reaction was then inactivated by incubation at 75 ºC for 10 minutes. The samples were centrifuged at 10,000 rpm at room temperature for 45 minutes and then 20 µl extracted in a 5:3:2 mixture of methanol:acetonitrile:sample, resulting in a final volume of 100 µl. After a final centrifugation at 10,000 rpm for 5 minutes, the upper 90 µl of the sample was taken and the organic solvent removed using a vacuum centrifuge (Thermo). Analytes were reconstituted in 30 µl mass spectrometry grade water by Jimi Wills and 10 µl injected onto a 30x1mm HyperCarb column (VWR). A gradient of 0-90% B was run over 4 minutes, where B is acetonitrile and A is 20 mM ammonium carbonate. Mass spectra were acquired in negative mode on a Thermo Q Exactive, scanning from 300 to 350 m/z at resolution 70k. AGC target was set to 1 x 10^6 and maximum ion time

100 ms. Data were analysed using an in-house R script, which is in re-submission and is available at https://gitlab.com/jimiwills/assay.R.

2.2.7 Bisulfite sequencing

The methylation status of genomic CpGs was examined using bisulfite sequencing. This involves bisulfite treatment of genomic DNA to distinguish methylated from unmethylated CpGs, as unmethylated cytosines are converted to uracil in the presence of bisulphite whereas methylated CpGs are protected. DNA Sanger sequencing can then be used then identify unmethylated (thymine) and methylated (cytosine) CpGs. EZ DNA Methylation-Lightening kit (Zymo) was used to bisulfite treat 500 ng genomic DNA using the manufacturer's instructions. Bisulfite treated and purified DNA was PCR amplified using primers specific to the region of interest (detailed in Table 2.3). The PCR was set up as follows:

2 µl bisulfite treated DNA

1 µl forward and reverse primer mix (10 µm of each primer)
2 µl 10x PCR buffer (Invitrogen)
2 µl dNTPs (10 µm of each dNTP, Invitrogen)
0.6 µl Mg2+ (Invitrogen)
0.2 µl Taq polymerase (Invitrogen)

The cycling conditions were as follows (35 cycles):

94 ºC 3 minutes
94 ºC 45 seconds
58 ºC 30 seconds
72 ºC 30 seconds
72 ºC 10 minutes

PCR products of the expected size (360 bp for major satellites and 210 bp for minor satellites) were gel purified, cloned into the pGEM-T Easy vector, transformed into DH5α competent cells (described in Section 2.5). Transformed bacterial colonies (white colonies) were picked and placed into wells of a 96-well mini-prep block containing 1 ml Luria-Bertani (LB) culture broth containing 50 µg/ml ampicillin. The 96-well blocks were incubated at 37 ºC overnight in a shaking incubator and plasmid DNA was mini-prepped and sequenced using Sp6 or T7 sequencing primers by the

HGU technical services. The resulting sequences were aligned to a reference sequence for the region and the methylation status of each CpG assessed using the Bisulfite Sequencing DNA Methylation Analysis (BISMA) online tool for repetitive sequences (http://services.ibc.uni-stuttgart.de/BDPC/BISMA/index_repetitive.php, Rohde *et al.*, 2010). Default settings were used except from the sequence identity which was set at 60%. Clonal sequences were identified by BISMA and removed. The overall CpG methylation percentage was calculated for each region. Bisulfite plots were generated using the online Quantitative Methylation Analysis tool (QUMA, http://quma.cdb.riken.jp/).

Primer name	Target	Primer sequence (5'-3')
Majsat_Bis	Major satellite repeats	F: GGAATATGGTAAGAAAATTGAAAATTATGG R: CCATATTCCAAATCCTTCAATATACATTTC
Minsat_Bis	Minor satellite repeats	F: TAGAATATATTAGATGAGTGAGTTATATTG R: ATTATAACTCATTAATATACACTATTCTAC

Table 2.3. Details of primers used for bisulfite sequencing.

2.3 RNA preparation, manipulation and analysis

2.3.1 RNA extraction from mammalian cells

Mammalian cell pellets that had been resuspended in 1 ml Trizol and stored at -80 ºC were thawed at room temperature. 200 µl chloroform was added to the sample followed by vigorous mixing by hand and incubation at room temperature for 5 minutes. Aqueous and organic phases were then separated by centrifugation at 13,000 rpm for 15 minutes at 4 ºC. The upper aqueous phase was removed and added to 1 volume of 70% ethanol. The remainder of the RNA extraction was then performed using the RNeasy mini kit (Qiagen) as per the manufacturer's instructions. Briefly, the aqueous phase/ethanol mixture was added to RNeasy extraction columns and centrifuged at 13,000 rpm for 30 seconds. The columns were then washed with the appropriate volume of RW1 wash buffer. An on-column DNase step was performed using RNase-free DNase (Qiagen), adding 10 µl DNase to 70 µl buffer RDD which was subsequently incubated on the column at room temperature for 15 minutes. The columns were washed again with buffer RW1, then washed twice with the appropriate

volume of RPE. A dry spin at 13,000 rpm for 1 minute was included to eliminate residual wash buffer. RNA was eluted in 30 µl of RNase-free water. RNA integrity was assessed by electrophoresis on a 1% agarose gel. 2x RNA loading dye (Thermo Fisher Scientific) was added to RNA samples which were then incubated at 80 ^{o}C for 2 minutes. RNA samples were loaded onto the gel alongside 2 µl RiboRuler (Thermo Fisher Scientific) to distinguish 28S and 18S RNA fragments. RNA was stored at -80 ^{o}C.

2.3.2 cDNA preparation

Complementary DNA (cDNA) was synthesised using Superscript III (Invitrogen) as per manufacturer's instructions. Initially, 2.5 µg total RNA was mixed with 100 ng random hexamers (Promega), 1 µl of 10 mM dNTP mix (Invitrogen) and distilled water up to a volume of 13 µl. This mixture was heated to 65 ^{o}C for 5 minutes then incubated on ice for at least 1 minute. The following components were then added resulting in a final volume of 20 µl:

4 µl 5x First-Strand buffer
1 µl 0.1 M DTT
1 µl RNaseIn ribonuclease inhibitor (Promega)
1 µl SuperScript III (200 units)

This reaction was incubated according to the following protocol:

25 ^{o}C for 5 minutes
50 ^{o}C for 1 hour
70 ^{o}C for 15 minutes

Negative control RT reactions were also set up as shown above without the addition of SuperScript III to control for the contribution of genomic DNA and aerosol contamination in downstream experiments. The resulting cDNA was diluted to 100 µl in distilled water and stored at -20 ^{o}C.

2.3.3 qRT-PCR

To determine the transcript levels from certain genes, qRT-PCR was performed using first-strand cDNA. Primers for qRT-PCR (detailed in Table 2.3) were designed using

the ‘Primer3’ program (http://frodo.wi.mit.edu/primer3/), with altered settings to give an annealing temperature of 60 ºC and a product size of < 300 bp. qRT-PCR primers were designed to span a large intron to avoid the detection of any contaminating genomic DNA. Reactions were set up with a final volume of 10 µl as follows:

5 µl 2x SYBR Select Master Mix (Thermo Fisher Scientific)
2 µl forward and reverse primer mix (2.5 µM of each primer)
1 µl cDNA
2 µl sterile water

The PCR was performed using a LightCycler480 (Roche) with the following cycling conditions:

95 ºC for 2 minutes (Initial denaturation)
95 ºC for 15 seconds (Denaturation)
60 ºC for 1 minute (Annealing and elongation)

Denaturation and annealing/elongation steps were repeated for 50 cycles in total.

Target gene	Primer sequence (5'-3')
Prdm14	F: TAATGCCAAAGCTCACCCGT R: GAGGCTCTTTGCCTCTCCAA
Nanog	F: TCGAATTCTGGGAACGCCTC R: GTCTTCAGAGGAAGGGCGAG
Oct4	F: GCCGTGAAGTTGGAGAAGGT R: CTTCTGCTTCAGCAGCTTGG
Dnmt3b	F: CTGGGTACAGTGGTTTGGTG R: AGTGTGGTACATGGCCTTCC
Dnmt3l	F: TTCTCACGGAGTGGACTGCT R: AAAAGGCTCAGTACCCGCA
Dnmt1	F: CTGGAAGAGGTAACAGCGGG R: CTGGTGTGACGTCGAAGACT
Fgf15	F: CGGTCGCTCTGAAGACGATT R: CCTCCGAGTAGCGAATCAGC
Pax6	F: CTTCAGTACCAGGGCAACCC R: GAGTTGGTGTTCTCTCCCCC
Klf4	F: AGAACAGCCACCCACACTTG R: AATTTCCACCCACAGCCGTC
Fgf5	F: AGTCAATGGCTCCCACGAAG R: GGCACTTGCATGGAGTTTTCC
Foxa2	F: GCCCGAGGGCTACTCTT R: CATTCCAGCGCCCACATAG
TBP	F: GAAGAACAATCCAGACTAGCAGCA R: CCTTATAGGGAACTTCACATCACAG
β-actin	F: AGAGCTATGAGCTGCCTGACG R: TGTGTTGGCATAGAGGTCTTTACG
Nestin	F: AGGACCAGGTGCTTGAGAGA R: TTCGAGAGATTCGAGGGAGA
Gata4	F: GGAAGACACCCCAATCTCGT R: ACATGGCCCCACAATTGACA
Brachury	F: TGGCTCTAGCCAGTATCCCA R: AAAGAACTGAGCTCCCAGCC

Table 2.4. Target genes and sequences of primers used for qRT-PCR analysis.

2.4 Protein preparation and analysis

2.4.1 Protein extraction

Whole cell extracts were prepared using Pierce RIPA (Thermo Fisher Scientific) buffer supplemented with 1X phosphatase inhibitor cocktail (Thermo Fisher

Scientific) and 1x protease inhibitor cocktail (cOmplete mini EDTA-free tablets, Sigma). To lyse the cells, 100 µl of supplemented RIPA buffer was added to approximately 1 x 10^6 frozen cells. This mixture was incubated on ice for 10 minutes. To degrade the nucleic acids in these extracts, 1 µl benzonase (Sigma) was added and the sample incubated for a further 30 minutes on ice. To enhance cell lysis, extracts were sonicated in a Bioruptor (Diagenode) for 10 cycles (30 seconds on/30 seconds off) on the high setting, then centrifuged for 10 minutes at 13,000 rpm at 4 ^{o}C to remove cell debris. The supernatant was taken for protein quantification. For storage, protein extracts were stored at -80 ^{o}C after adding glycerol to a final concentration of 10%.

2.4.2 Protein quantification

Protein concentration of whole cell extracts was measured using the Pierce BCA protein assay kit (Thermo Fisher Scientific) as directed by the manufacturer's instructions. Samples were diluted 1 in 4 so that the protein concentration fell within the range of the standard curve in the assay. Each sample was measured in duplicate in a 96-well flat-bottomed plate and the protein concentration read in a Multiskan GO microplate spectrophotometer (Thermo Fisher Scientific) at 562 nm. Protein concentrations were calculated using the spectrophotometer readings from the standard curve, taking into account the dilution factors of the extracts. Between 20 and 40 µg of protein was used for Western blotting analysis.

2.4.3 SDS-PAGE

Before loading protein samples onto a gel, 4x NuPAGE sample loading buffer and 10x NuPAGE sample reducing agent (Invitrogen) were added to the appropriate volume of cell extract to give a final concentration of 1x. Samples were boiled at 70 ^{o}C for 10 minutes then incubated on ice prior to loading. Protein was resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using NuPAGE 4-12% Bis-Tris precast protein gels (Invitrogen). 4 µl PageRuler prestained protein ladder (Invitrogen) was also loaded for sizing. Gels were run in a NuPAGE gel tank (Invitrogen) in 1x NuPAGE MOPS SDS running buffer (Invitrogen) at approximately 150 V for 1-1.5 hours.

2.4.4 Western blotting

Following protein separation by SDS-PAGE the proteins were transferred onto a nitrocellulose membrane using the iBlot 2 dry transfer system (Invitrogen), The P0 transfer method was used, which involves transfer at 20 V for 1 minute, 23 V for 4 minutes and 25 V for 2 minutes. Membranes were stained with Ponceau S solution (Sigma) following transfer to assess whether equal transfer had occurred. Blocking was performed for approximately 1 hour using 1.25 ml Western blocking reagent (Roche), 1x Tris buffered saline (TBS, 50 mM Tris, 150 mM sodium chloride, pH adjusted to 7.5) made up to 50 ml with sterile water. Membranes were incubated overnight at 4 ^{o}C with the appropriate primary antibody (α-HELLS, Proteintech, 11955-1-AP, 1:1000; α-β-TUBULIN, Abcam, ab6046, 1:1000) in block solution. Following overnight incubation, blots were washed with 1x TBS supplemented with 0.1% Tween-20 (TBST, Sigma) 3 x 5 minutes at room temperature on a shaking platform. Secondary antibodies (α-mouse or α-rabbit HRP-conjugate, Cell signalling, 1:10,000) were incubated with the membranes in block solution for 1-2 hours at room temperature, then thoroughly washed 3 x 10 minutes in TBST. Protein detection was performed using SuperSignal Western Pico Reagent chemi-luminescence reagent (Pierce) for 5 minutes, then developed using ECL Hyperfilm (Amersham) and an x-ray film processor.

2.5 Cloning techniques

2.5.1 PCR amplification

To amplify fragments for TA cloning, reactions were set up for PCR amplification as follows:

25 µl DreamTaq green 2x PCR master mix (Thermo Fisher Scientific)
3 µl forward and reverse primer mix (10 µm of each primer)
250 ng template DNA
Up to 50 µl with distilled water

Cycling conditions were as follows (35 cycles):

95 ^{o}C 3 minutes
95 ^{o}C 30 seconds

T_a 30 seconds
72 ^oC 1 minute
72 ^oC 10 minutes

The primers and annealing temperatures (T_a) used are indicated in Table 2.4.

2.5.2 DNA fragment extraction from agarose

Amplified DNA fragments were resolved using 1% agarose-TAE gels supplemented with 1x SYBR safe and were viewed using a Safe Imager transilluminator (Invitrogen). DNA fragments were purified from agarose gels using the Nucleospin PCR and gel extraction kit (Machery-Nagal) as per the manufacturer's instructions. The DNA fragment(s) of interest were cut from the gel using a scalpel blade. Gel slices were weighed, 200 µl buffer NT1 per 100 mg agarose was added and then incubated at 50 ^oC for 10 minutes in a Thermomixer with shaking at 1000 rpm to dissolve the gel slices. The sample was centrifuged at 11,000 rpm through the silica-gel columns included in the kit for 1 minute. DNA bound to the column was washed twice with buffer NT3 supplemented with 100% ethanol then dried by centrifugation for 1 minute at 11,000 rpm. DNA was eluted from the columns in 15 µl elution buffer by centrifugation for 1 minute at 11,000 rpm following a 1 minute incubation at room temperature.

2.5.3 pGEM T Easy cloning

DNA fragments amplified by *Taq* polymerase often have a 3' deoxyadenosine (dA) which means they can be efficiently ligated into vectors with 5' deoxythymidine (dT) overhangs. Purified DNA fragments were therefore ligated into the linearised pGEM T Easy vector (Promega) which contains 5' dT overhangs. Ligation was performed using 50 ng pGEM T Easy vector and the Rapid DNA ligation kit (Roche), with an approximate insert to vector ratio of 3:1. Ligation reactions were incubated at room temperature for at least an hour or overnight at 4 ^oC.

2.5.4 Bacterial transformation

Library efficiency DH5α chemically competent *Escherichia coli* cells (Invitrogen) were used for bacterial transformation. 50 µl of competent cells were mixed with 1 µl

ligated DNA and incubated on ice for 20-30 minutes. The transformation mixture was heat shocked at 42 ^{o}C for 45 seconds then immediately cooled on ice for at least 2 minutes. Room temperature S.O.C media was added to the transformations up to a volume of 500 µl. The transformations were subsequently incubated at 37 ^{o}C in a shaking incubator for 1 hour, then 1/10th and 9/10th of the mixtures were aseptically spread onto separate L-agar plates containing 50 µg/ml ampicillin and 200 µg/ml X-gal (Sigma) for blue-white colony selection. The plates were inverted and incubated at 37 ^{o}C overnight to allow colonies of transformed bacteria to grow.

2.5.5 Bacterial culture and isolation of plasmid DNA

For the preparation of plasmid DNA, single white (indicating that they were recombinant) colonies were picked and added to 5 ml LB broth containing 50 µg/ml ampicillin. They were then incubated at 37 ^{o}C overnight in a shaking incubator set at 225 rpm. Plasmid DNA was extracted using the NucleoSpin Plasmid kit (Machery-Nagal) as per the manufacturer's instructions, eluting in 50 µl of the buffer AE provided.

2.5.6 Sanger sequencing

Isolated plasmid DNA was sequenced by the HGU technical services using Sp6 or T7 sequencing primers. The BigDye terminator v3.1 sequencing kit was used according to manufacturer's instructions and products run on an ABI Prism 3720 genetic analyser. Sequence data was analysed with Bioedit Sequence Alignment Editor version 7.2.5.

Primer name	Target	Primer sequence (5'-3')	T_a (ºC)	PCR product (bp)
Sequence mutation checks				
HELLS_EX2_4	*Lsh* exon 1	F: TTGAATCCCGTTCTGAAAGC R: GGCCGTGGACTTTACATTACC	58	182
3L_del	*Dnmt3l* exon 1	F: AGACCCTTGAAACCTTGGAC R: TACCTGGTAGGAGAGGGTGC	58	303
HELLS_HAP1	Human HELLS exon 3	F: GTTCCAGTTTTTGGTTTGAAAGTGT R: ATGGAAGTGCACTGTATCTTGATTG	53	434
CRISPR off-target mutation checks				
Chr11:97a	*Tbkbp1*	F: TCGGGTTGGAATGCACGTAG R: GCTTCAGCGCTTCCTGTTTC	60	304
Chr11:107a	Non-genic	F: GTCCAGCCTGGCTTCACAG R: CACACCCCACCCAAGATACC	60	306
Chr16:98a	Non-genic	F: ATGCAGGAACAGAGCCCATC R: CCTATAGAGCTAGAGTGGCTGG	60	299
Chr2:155a	*Myh7b*	F: TCCTGCCCTCTCTGGATCAG R: GCCCTAGCCTAAGGAGATGG	60	301
Lsh blastocyst genotyping				
WT	*Lsh* intron 12	F: TGTGGATAATTCCATTTCACAG R: ACAAGTTTGTACAAAAAAGCAGGCT	55	600
KO	Genetrap cassette	F: TCATAGAAGGAAACCTGCATAGAAG R: TCCAAACTCATCAATGTATC	55	260

Table 2.5. Details of primers used for sequence mutation checks, CRISPR off-target mutation checks and Lsh blastocyst genotyping.

2.6 Mammalian cell imaging

2.6.1 Immunofluorescence

Cells were plated onto glass coverslips coated in 0.2% gelatin in 12-well culture plates at a density of approximately 1-2 x 10^5 cells per well and cultured overnight. Media was removed and cells washed with PBS before adding 0.5 ml 4% paraformaldehyde (PFA, w/v, Sigma) per well for 10 minutes at room temperature for fixation. PFA was removed and fixed cells washed twice with PBS before being blocked with 10% donkey serum in 0.1% Triton-X (v/v, both Sigma) for at least 1 hour at room temperature. Following blocking, primary antibodies (detailed in Table 2.5) were

added to the cells in 400 µl total volume of 1% donkey serum in 0.1% Triton-X and incubated at 4 ^{o}C overnight. Primary antibody solution was washed off with 3 x 10 minutes washes in PBS at room temperature. Secondary antibodies (α-mouse or α-goat AlexaFluor-555 or -647, Thermo Fisher Scientific, 1:200) were added to cells in the same blocking solution as primary antibodies for 45-60 minutes at room temperature in the dark. Cells were washed again in PBS for 3 x 10 minutes in the dark at room temperature. DAPI (50 µg/ml, 4',6-diamidino-2-phenylindole) was diluted 1 in 1000 in PBS then 500 µl added per well for 10 minutes at room temperature in the dark. Coverslips were washed 2 x 10 minutes in PBS and once in distilled water then mounted onto glass slides using Vectashield (Vector Laboratories) and sealed using nail varnish. Secondary-only controls where the primary antibody incubation was omitted were also performed assess the presence of any background signal. Images were captured using the 40x objective of a Zeiss Axioplan II fluorescence microscope and micromanager open source microscopy software. Images were analysed using ImageJ software. Signal correlation plots were generated using the Coloc2 plugin using DAPI signal to define a region of interest for analysis.

Antibody	Catalogue number	Company	Dilution used	Secondary antibody
Nanog	AF2729	R&D systems	1:75	Goat
Esrrb	pp-h6705-00	R&D systems	1:500	Mouse
Oct4	Sc-5279	Santa Cruz Biotechnology	1:400	Mouse
α-smooth muscle actin	ab7817	Abcam	1:100	Mouse
Tuj1	ab14545	Abcam	1:1000	Mouse

Table 2.6. Details of primary antibodies used for immunofluorescence.

2.6.2 Brightfield imaging

Phase contrast microscopy was used to capture brightfield images of live mammalian cells in culture. The 10x objective of a Nikon Eclipse Ti-S inverted microscope was used to capture images along with iVision 4.5 software.

2.7 Generation and bioinformatic analysis of next-generation sequencing data

2.7.1 Transcripome analysis by RNA-seq

2.7.1.1 Library preparation and sequencing

Genome-wide expression analysis by RNA-seq was performed to analyse the transcriptomes of cultured cells. The integrity of the starting total RNA was assessed using a Bioanalyzer 6000 nano chip (Agilent, run by HGU technical services). Only high quality RNA with a RIN score > 8 was used for downstream analyses.

For serum and 2i mESCs, library preparation was performed in-house using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) as per manufacturer's instructions to purify polyadenylated RNA from 1 µg total RNA. Libraries were then generated from the poly(A)-purified material using the NEBNext Ultra Directional RNA library prep kit for Illumina (NEB) and NEBNext Adaptors for Illumina (NEB) as per manufacturer's instructions. During library preparation, PCR library enrichment was performed for 13 cycles. Agencourt AMPure XP beads were used throughout to purify samples following each reaction. The quality of the resulting libraries were assessed on a Bioanlyzer DNA high sensitivity chip (Agilent, run by HGU technical services). This confirmed the absence of smaller fragments representing primers and adaptor-dimers. For EpiSCs, TruSeq stranded mRNA-seq libraries were prepared by Edinburgh Genomics. All libraries were sequenced using Illumina HiSeq 4000 75 bp paired-end read sequencing performed by Edinburgh Genomics.

2.7.1.2 Mapping and normalisation of RNA-seq data

All RNA-seq datasets were downloaded from Edinburgh Genomics and tested for sequencing read quality using FastQC (v0.11.4). Once data quality was confirmed, demultiplexed datasets were parsed for NGS adaptor presence, short reads, read quality scores and reads falling below our thresholds were filtered out using Trim Galore! (v0.4.1.). Paired processed fastq files (Paired End sequencing was employed) were mapped to the mouse genome (genome build Mus_musculus.NCBIM37.61.fa and the matched transcriptome assembly Mus_musculus.NCBIM37.61.gtf) using

TopHat (v2.1.0) using the settings: tophat --no-coverage-search -r {inner mate distance} --mate-std-dev {mate distance standard deviation} --library-type fr-firststrand -G /path/to/transcriptome/assembly -p 8 –o output.file /path/to/genome/build a.fastq b.fastq. Inner mate distances and standard deviations were computed for each paired set of fastq files using an unpublished python script kindly provided by Prof. Ian Adams. Counts over transcripts were counted using HTSeq count tool (v0.7.1) and were converted to an edgeR (v3.12.1) objects in the R (v3.4.0) programming environment using normalisation based on count depths over given datasets with limma (v3.26.9).

2.7.2 DNA methylome analysis by enhanced reduced representation bisulfite sequencing (ERRBS)

Genome-wide DNA methylation profiles were generated for serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs using ERRBS. The integrity of the starting RNase-treated genomic DNA was assessed by gel electrophoresis and nanodrop 260/280 and 260/230 readings prior to being sent for library preparation. 500 ng of genomic DNA at 20 ng/µl was sent to the Epigenomics Core Facility of Weill Cornell Medicine (New York, USA) for library preparation using the protocol described in Garrett-Bakelman *et al.* (2015).

2.7.2.1 Mapping and normalisation of ERRBS data

ERRBS was carried out by the Epigenomics Core Facility of Weill Cornell Medicine (New York, USA) who provided custom data analysis. In addition, from our background in NGS analysis ‘in-house’, we decided to analyse the ERRBS data de novo. To avoid batch effects on the sequencing instrument, libraries were run over three individual lanes per flow cell. In Unix, these ‘triplicates’ were joined by catenation, and analysed as one entity. As for RNA-seq, various quality control metrics were assessed prior to full dataset mapping, including Trim Galore!. Data were mapped to a mouse genome build (Mus_musculus.NCBIM37.61.fa) using the Bismark tool suite (v0.16.3) with the following settings: bismark --bowtie2 /path/to/mm9/genome/index/ -n 1 -l 32 -non_directional --un --ambiguous -o ./output_pe --non_bs_mm -1 /path/to/1.fq.gz -2 ./path/to/2.fq.gz. Next,

‘bismark_methylation_extractor’ was used to extract methylation scores over CpG dinucleotides from the bismark output above. [Note: the methylation scores were also extracted using custom scripts by Dr Dunican (unpublished) which gave identical results]. Outputs were filtered for a minimum read-depth of 5 reads per CpG location. [Reads-depths of 10+ gave identical data trends, with less data]. All libraries yielded comparable read counts, so inter-library normalisation was deemed unnecessary. Changes in methylation were computed for the various cell lines concerned and smoothScatters, violin plots (vioplot v0.2) and pie-charts were produced in R (v3.4.0).

Chapter 3. *Lsh* is not required to facilitate establishment of DNA methylation during reversion from the ground state of pluripotency

3.1 Introduction

In this chapter I discuss the optimisation of a culture system in which to dynamically modulate DNA methylation in mESCs. As outlined earlier, conventional mESC culture in which serum is included in the media (hereafter referred to as "serum" mESCs) results in a population of cells displaying heterogeneous colony morphology and expression of key pluripotency factors. Transferring mESCs into media containing two small molecule kinase inhibitors (hereafter referred to as "2i" mESCs) transitions the cells into a naïve "ground state" of pluripotency in which they more closely resemble cells of the ICM *in vivo* (Ficz *et al.*, 2013; Habibi *et al.*, 2013; Choi *et al.*, 2017). This transition of cell state coincides with more homogeneous colony morphology, gene expression changes and global hypomethylation. There is published evidence to suggest that these two culture states are interconvertible, and that the global hypomethylation observed in 2i mESCs can be reversed (Habibi *et al.*, 2013; Leitch *et al.*, 2013). By harnessing the interconvertibility of this system, I aimed to dynamically modulate global DNA methylation levels by altering mESC culture conditions. Specifically, I intended to use this experimental set-up to determine the requirement for the putative chromatin remodelling factor *Lsh* in re-methylation of the genome during the transition from 2i to serum culture. I was also able to adapt this culture system to investigate the involvement of other factors in re-establishment of DNA methylation during serum reversion, such as DNA methyltransferase co-factor *Dnmt3l*.

At the level of DNA methylation, 2i mESCs have been proposed to be more representative of cells of the ICM, whereas DNA methylation levels of serum mESCs are more similar to that of differentiated cells (Habibi *et al.*, 2013; Choi *et al.*, 2017). Therefore, it can be reasoned that the re-establishment of DNA methylation that occurs during reversion from 2i could model the *de novo* methylation that occurs in the early

stages of implantation of the blastocyst *in vivo*. Therefore, using this culture system I could infer the involvement of *Lsh,* and other factors, in DNA methylation deposition during the early stages of embryonic development.

3.2 Results

3.2.1 Characterisation of lsh$^{-/-}$ mESCs

As discussed in the introduction, extensive analysis has been undertaken in *lsh*$^{-/-}$ embryonic tissues and somatic cell lines. However, analysis of the impact of *Lsh* deletion in a pluripotent context has been more limited, despite *Lsh* expression being comparatively much higher in pluripotent cells (Raabe *et al.*, 2001; Xi *et al.*, 2009). Therefore, I decided to use mESCs as a model to investigate the function of *Lsh* in early embryonic development. Prior to me joining the group, Dr Donncha Dunican used CRISPR/Cas9 gene editing technology to generate *lsh*$^{-/-}$ mESCs. A single-guide RNA (sgRNA) was designed to target exon 1 of *Lsh*. This was transfected into E14 mESCs along with WT Cas9. Clones were screened by *XhoI* restriction digest followed by PCR to identify mESC lines with a mutation in *Lsh*. The nature of the mutation was then assessed by Sanger sequencing. One clone was identified as having a 25 bp homozygous deletion in *Lsh* resulting in a premature stop codon in exon 1. This clone was subsequently shown to have no LSH protein expression by Western blot. I verified the homozygous mutation in this clone by Sanger sequencing and confirmed the lack of LSH protein expression by Western blot (Figure 3.1 A and B respectively). Confident that there was a homozygous deletion in this clone resulting in ablation of LSH protein levels, I used this *lsh*$^{-/-}$ mESC clone and the WT E14 counterpart for all downstream analyses.

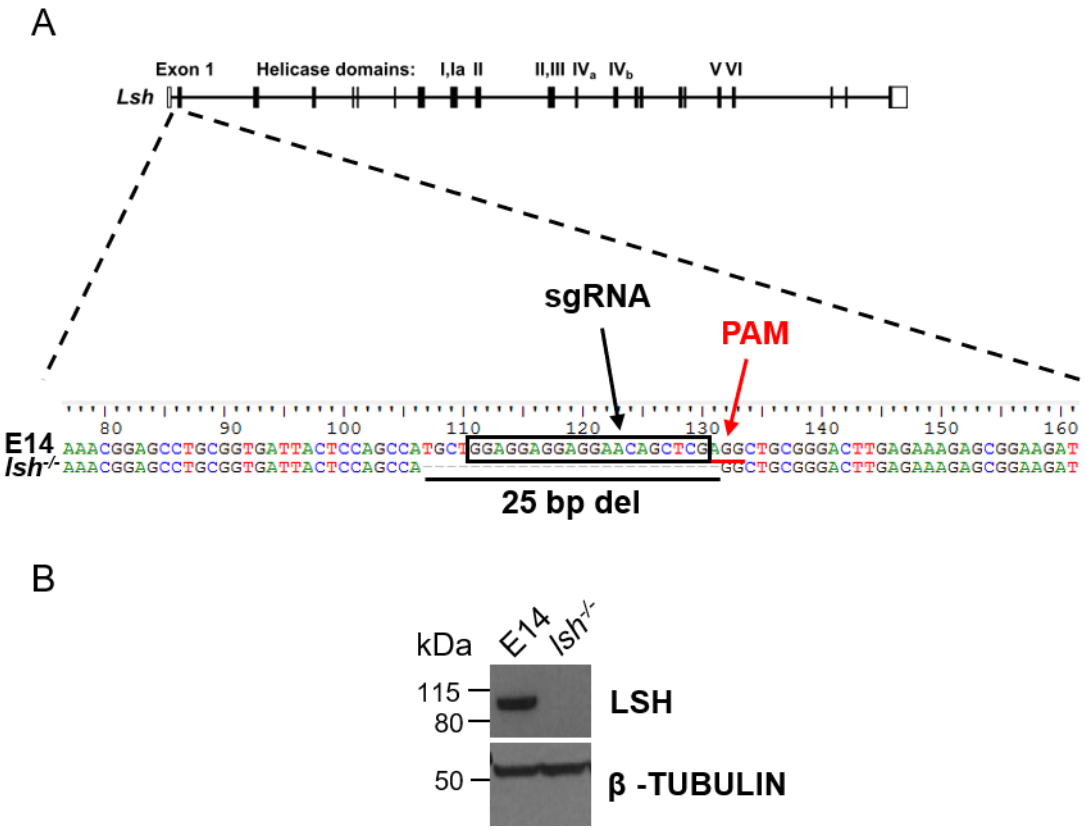


Figure 3.1. Confirmation of CRISPR-generated *lsh*$^{-/-}$ mESCs. A. *Lsh* exon 1 was targeted for CRISPR-mediated gene editing, which was assessed using Sanger sequencing of a region encompassing the sgRNA target site and protospacer adjacent motif (PAM) site, which are highlighted in the E14 sequence. The sgRNA determines target specificity of the Cas9 nuclease during the CRISPR gene editing process, whereas the PAM site is necessary for Cas9 binding and cleavage of the target DNA sequence. The 25 bp deletion present in *lsh*$^{-/-}$ mESCs is also highlighted. B. Western blot analysis of whole protein extracts from E14 and *lsh*$^{-/-}$ mESCs using an antibody against HELLS (human homolog of LSH) to determine LSH protein levels. An antibody against β-TUBULIN was used as a loading control.

3.2.2 Derivation of mESCs from lsh$^{-/-}$ mouse blastocysts

To independently verify results generated from the CRISPR-targeted *lsh*$^{-/-}$ mESCs, I aimed to derive *lsh*$^{-/-}$ mESC lines from an *Lsh* conditional knockout (KO) mouse model. This mouse model was generated by Dr Ian Adams and David Read using a genetrap construct obtained from the European Conditional Mouse Mutagenesis Program (EUCOMM, construct detailed in Figure 3.2 A). The ICM of the blastocyst was isolated by Dr Ian Adams from embryos produced by crossing two mice heterozygous for this conditional KO construct. The resulting mESC lines were derived in 2i medium on a layer of mitotically inactivated CD1 MEF feeders, following

the protocol described by Czechanski *et al.* (2014). This led to the derivation of 21 mESC lines. PCR genotyping was performed using the primers indicated in Figure 3.2 A and revealed that three mESC lines were homozygous for the conditional KO cassette, nine were heterozygous and nine were WT. An example of the PCR genotyping is shown in Figure 3.2 B. LSH protein levels were assessed in selected WT, heterozygous and homozygous mESC clones by Western blot. This revealed that there was residual LSH protein expression in clones shown to be homozygous for the genetrap cassette by PCR (Figure 3.2 C). This can be explained by so-called 'leaky expression', where the splice acceptor site in the genetrap cassette is skipped, resulting in full length protein production. This issue can be overcome by expressing *Cre* recombinase in the homozygous mESC lines to remove the critical exon 12 which is spanned by loxP sites. This, in theory, would prevent the production of any functional LSH protein, creating a complete *Lsh* knockout mESC line. However, this procedure could not be undertaken, as the blastocyst-derived mESC lines exhibited unstable growth following their recovery from cryopreservation. These mESCs grew slowly, with a high occurrence of cell death, and did not survive in long-term cultures. This was apparent for all the mESC lines generated during this derivation experiment, irrespective of the *Lsh* genotype; WT, heterozygous and homozygous mESC lines all exhibited unstable growth following their recovery from cryopreservation. This demonstrated that the absence of LSH protein did not contribute to the poor growth of these mESC lines in culture.

The inability to maintain these blastocyst-derived mESC lines in long-term cultures meant that it was not practical to continue to use these mESCs for this study, as the downstream experiments required cultures to be maintained for long time periods. Therefore, all future experiments were carried out using E14 and CRISPR-generated $lsh^{-/-}$ mESCs.

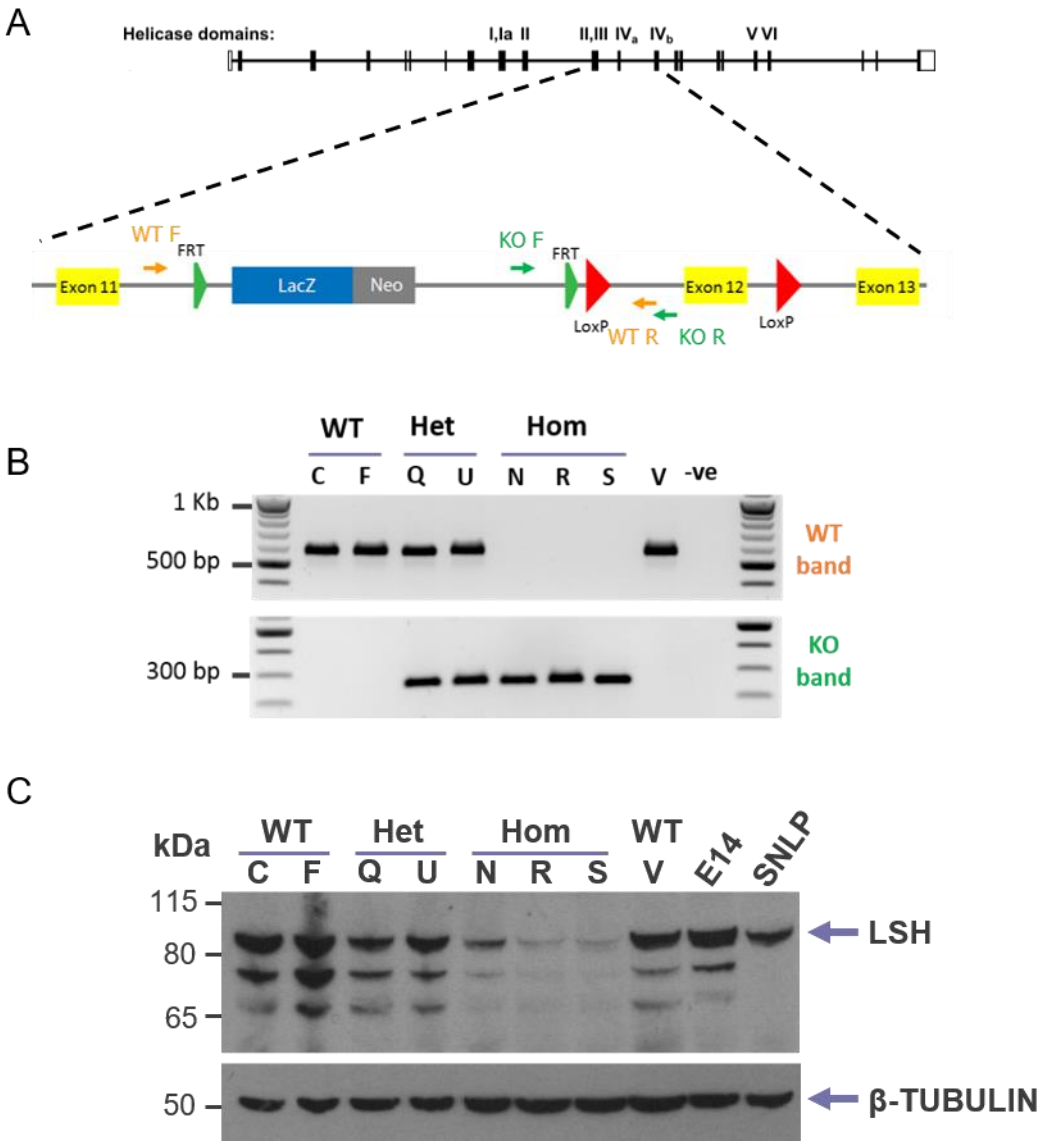


Figure 3.2. Derivation of *lsh*$^{-/-}$ mESCs from mouse blastocysts. A. Schematic showing design of the conditional knockout genetrap cassette received from EUCOMM. The genetrap cassette sits between *Lsh* exons 11 and 12 which correspond to helicase domains II, III and IV. The location of the genotyping primers are indicated by the orange arrows for the WT band and green arrows for the KO band. B. Agarose gel showing genotyping by PCR of selected mESC clones derived from *Lsh* conditional knockout mice. C. Western blot of whole protein extracts from mESC lines shown in B. An antibody against HELLS was used to determine LSH protein levels, β-TUBULIN was used as a loading control. Protein Extracts from E14 mESCS and SNLP feeders on which the blastocyst-derived mESCs were grown were included for comparison.

3.2.3 Optimisation of a culture system to modulate global DNA methylation levels in mESCs

3.2.3.1 Experimental strategy

To explore the requirement for *Lsh* in re-methylating the genome during reversion from 2i to serum culture, I aimed to optimise and validate a culture system using these culture conditions to dynamically modulate global DNA methylation. By harnessing the interconvertibility between these culture states, I hypothesised that I could reduce genome-wide DNA methylation levels by adapting mESCs to 2i culture in the presence of LIF, then return these mESCs to serum media to recover global DNA methylation levels (illustrated in Figure 3.3). As reported by Habibi *et al.* (2013), mESC culture in 2i media (with LIF) for at least 12 days is required to reduce global DNA methylation down to steady-state levels. Similarly, a reversion period in serum culture of at least 10 days appears to be required for re-establishment of global DNA methylation to the levels observed prior to 2i adaptation. Therefore, it was reasoned that a period of 15 days (six passages) for both adaptation to 2i and reversion into serum would be sufficient to induce genome-wide hypomethylation and subsequent re-methylation. To assess whether transition between the distinct culture states had occurred, a number of factors and metrics would be examined, including colony morphology and expression of key pluripotency/lineage-associated genes (highlighted in Figure 3.3).

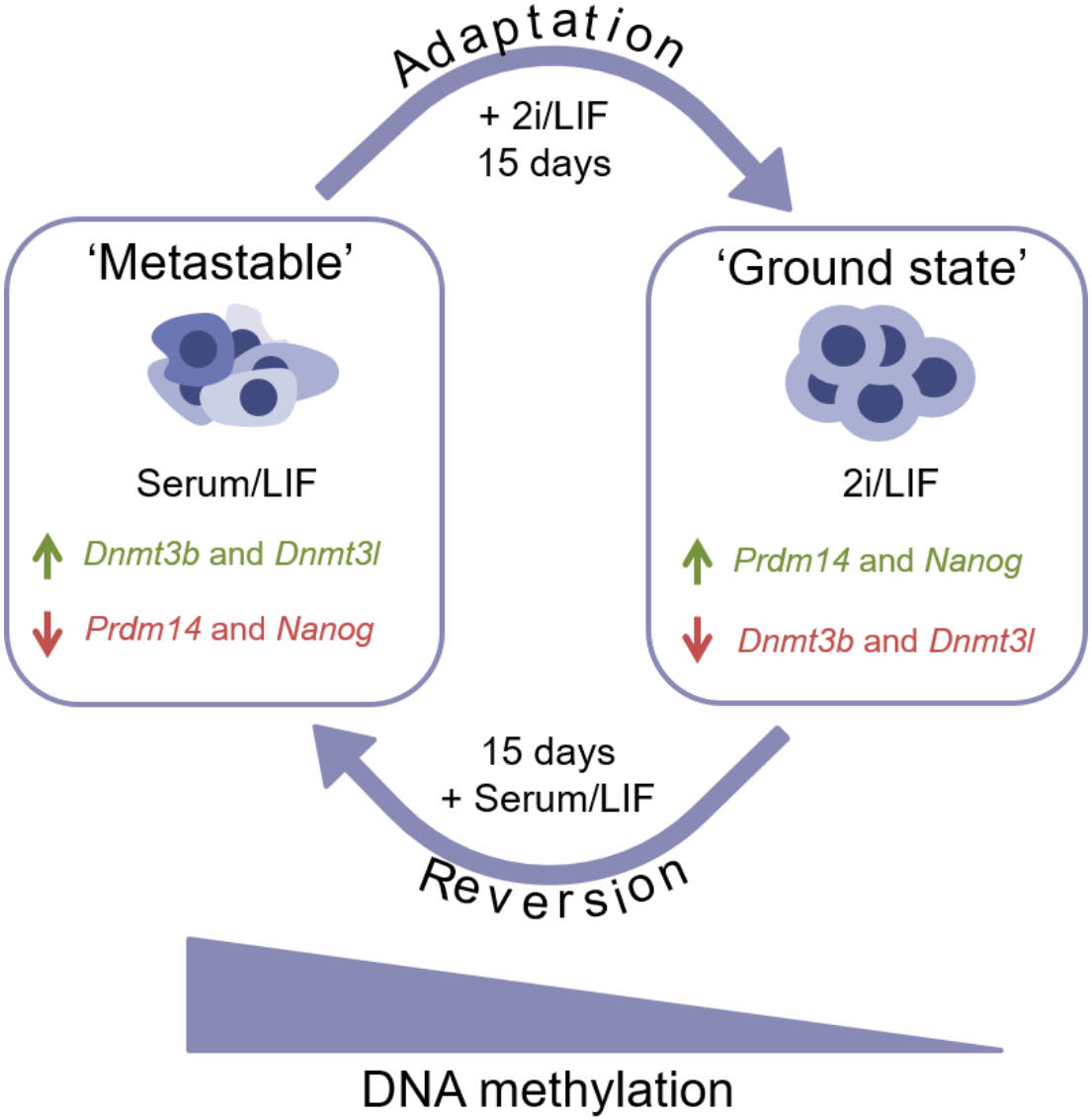


Figure 3.3. Schematic depicting 2i adaptation and serum reversion experimental design. Key genes that show expression differences between culture states are highlighted, as is the change in global DNA methylation levels. mESCs were cultured in either serum/LIF or 2i/LIF for 15 days (six passages) to induce transition between these distinct culture states.

3.2.3.2 E14 and lsh$^{-/-}$ mESCs change morphologically upon adaptation to 2i and reversion into serum

Morphological changes are a straightforward way to quickly assess the response of mESCs to the addition of 2i media. During adaptation to 2i, mESCs adopt a more uniform colony morphology, with individual colonies becoming more rounded and three-dimensional with smooth, defined colony boundaries (example shown in Marks

and Stunnenberg, 2014). This contrasts with the appearance of serum mESCs, which display a wide variety of colony morphologies, with a more flattened appearance and irregular colony edges. Both E14 and *lsh*$^{-/-}$ mESCs show this heterogeneous colony morphology whilst grown in serum (Figure 3.4, left panels). After adaptation to 2i for eight days (three passages), E14 mESCs had adopted the features of 'typical' 2i colony morphology described above (indicated by the arrow in Figure 3.4, middle panels). The appearance of *lsh*$^{-/-}$ mESCs following 2i adaptation for eight days (three passages) differed considerably compared to their serum counterparts. In comparison to the E14 mESCs, *lsh*$^{-/-}$ mESC 2i colonies displayed many features of typical 2i morphology such as becoming more rounded and three-dimensional. However, the colony boundaries did not become as smooth and defined as in E14 2i colonies, with some colonies exhibiting small protrusions or patches of flattened cells (indicated by the arrows in Figure 3.4, middle panels). Overall, although both E14 and *lsh*$^{-/-}$ mESCs display substantial changes in colony morphology upon adaptation to 2i, *lsh*$^{-/-}$ mESCs did not achieve the changes akin to typical 2i colony morphology to the same extent as E14 mESCs. These differences in the extent of colony morphology adaptation persisted to the final 15-day timepoint of the conversion to 2i (data not shown).

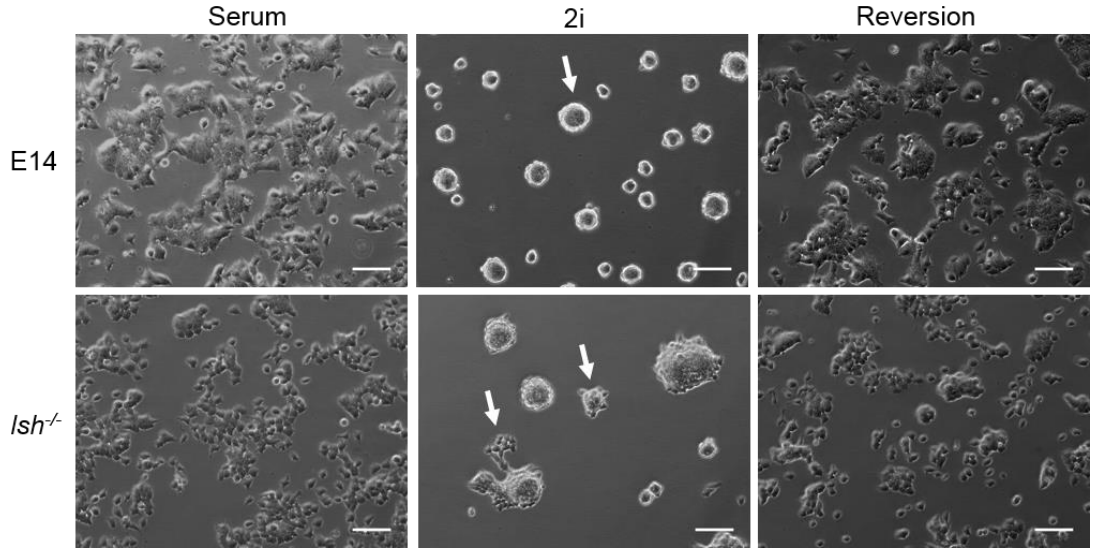


Figure 3.4. E14 and *lsh*$^{-/-}$ mESCs show similar morphological changes after longer term 2i adaptation and reversion back into serum culture. E14 and *lsh*$^{-/-}$ mESCs cultured in media containing serum (left panels), adapted to 2i for 8 days (3 passages, middle panels) and reverted into media containing serum for 9 days (4 passages) following 15 days of 2i adaptation (right panels). Arrows indicate typical colony morphology of each mESC line in 2i media. Scale bars represent 100 µm.

Following adaptation to 2i for 15 days, both E14 and *lsh*$^{-/-}$ mESCs were reverted into serum culture for 15 days. After nine days (four passages) of reversion, the morphology of both E14 and *lsh*$^{-/-}$ mESCs returned to the heterogeneous colony morphology observed prior to 2i adaptation (Figure 3.4, right panels). This demonstrates that both E14 and *lsh*$^{-/-}$ mESCs undergo dynamic and extensive morphological changes during the 2i adaptation and serum reversion process. However, morphological changes can only be considered as indicative of adaptation to 2i or serum culture. Further investigation into the expression of key genes is required to confirm successful transition between the cell culture states.

3.2.3.3 Nanog and Essrb expression becomes more homogeneous upon adaptation to 2i

Homogeneous *Nanog* expression is associated with transition to 2i and is often used as a marker to indicate adaptation to 2i culture. Immunofluorescence was performed on E14 and *lsh*$^{-/-}$ mESCs cultured in serum, adapted to 2i and reverted into serum to assess the uniformity of expression of *Nanog* and *Esrrb*, a downstream target of *Nanog*. In serum E14 mESCs, both *Nanog* and *Esrrb* appeared to be heterogeneously expressed, with some cells in the colonies exhibiting high *Nanog* and *Esrrb* expression and other cells displaying low expression. Upon adaptation to 2i, these factors became more uniformly expressed in all cells of the colony. When these 2i E14 mESCs were reverted into serum culture, *Nanog* and *Esrrb* expression also reverted to more heterogeneous expression (Figure 3.5 A and B, top panels). The same changes in expression patterns are evident in serum, 2i and reversion *lsh*$^{-/-}$ mESCs, however the increase in homogeneity of *Nanog* expression is not as obvious as in E14 mESCs, as *Nanog* expression already appears to be more uniform in serum *lsh*$^{-/-}$ mESCs (Figure 3.5 A and B, bottom panels). The relevant secondary antibody-only control staining for these experiments is presented in Appendix Figure A1. A and B.

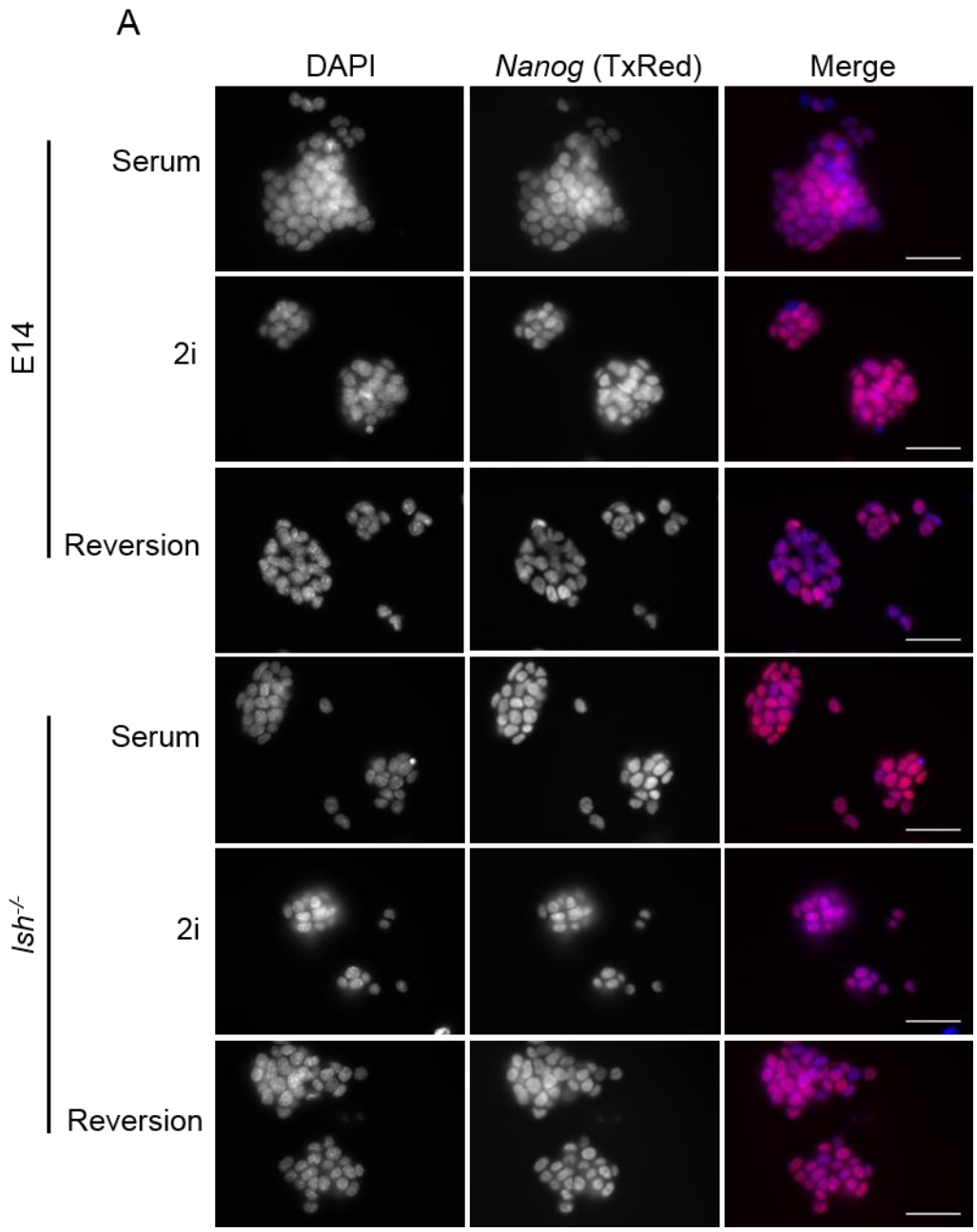
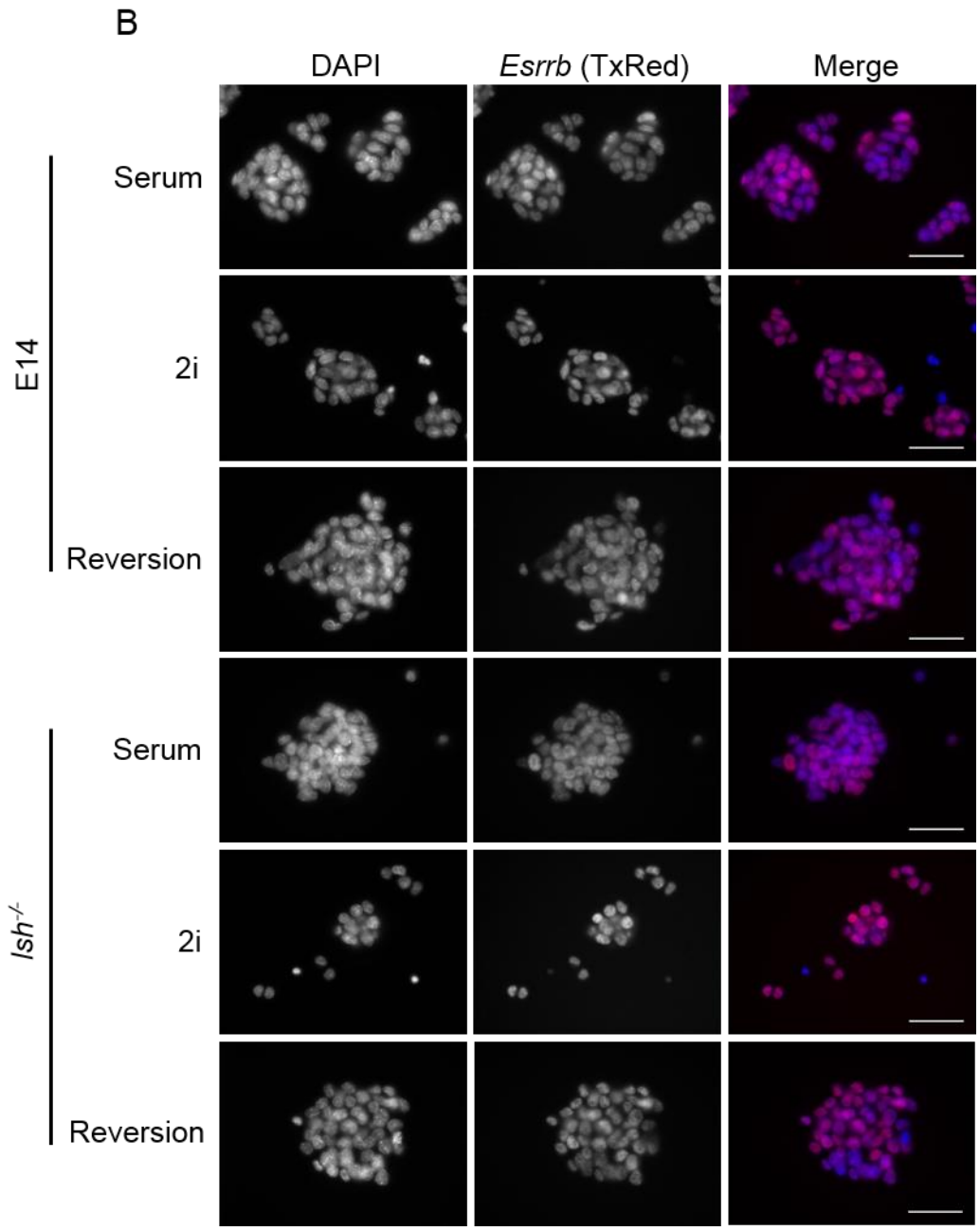


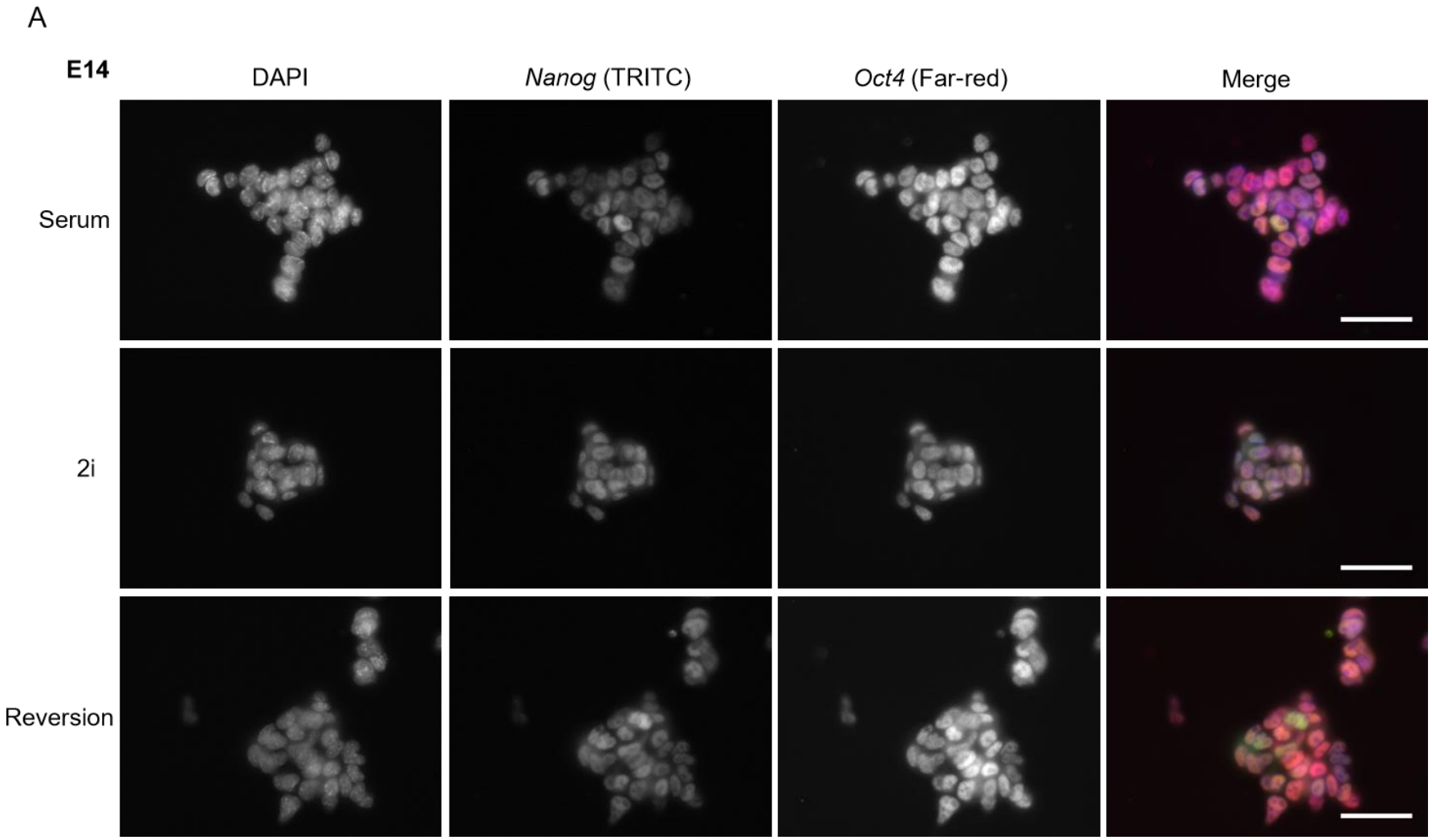
Figure 3.5. *Nanog* and *Esrrb* expression becomes more homogeneous upon adaptation to 2i and reverts to heterogeneous expression following reversion into serum in both E14 and *Ish*$^{-/-}$ mESCs. Immunofluorescence on E14 and *Ish*$^{-/-}$ mESCs cultured in serum, adapted to 2i for 15 days and reverted into serum culture for 15 days following 2i adaptation, using an antibody against *Nanog* (A) or *Esrrb* (B). Scale bars represent 50 µm.

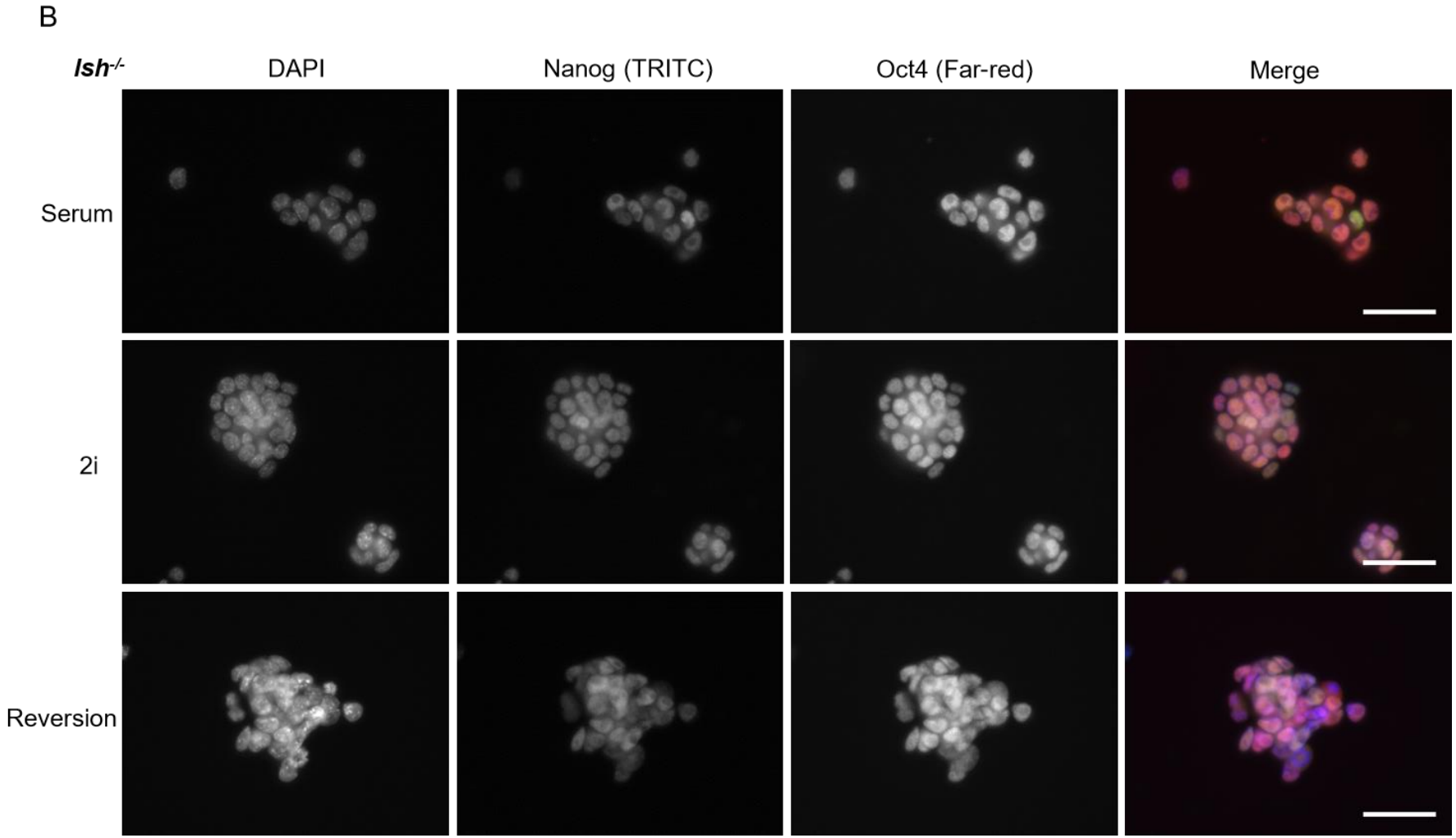


The difference in uniformity of *Nanog* expression between serum and 2i culture is more apparent when compared to a factor that is expressed homogeneously in both culture states. Therefore, double immunofluorescence was performed to simultaneously stain *Nanog* and *Oct4* in serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs. The secondary antibody-only control staining relevant to this double immunofluorescence analysis is presented in Appendix Figure A1. C. As observed previously, *Nanog* exhibited heterogeneous expression in serum E14 mESCs (Figure 3.6 A). Expression of *Oct4* however was more consistent, and when merged with DAPI

and *Nanog* signal it highlighted the heterogeneity of *Nanog* expression. Again, *Nanog* expression became more homogeneous following culture in 2i media, whereas *Oct4* expression remained uniform. This was again emphasised by the merged image, as most cells appeared to be similar in colour due to both factors being more uniformly expressed. After reversion into serum, *Nanog* returned to a heterogeneous expression pattern similar to that observed initially in serum. *Oct4* expression continued to appear more homogeneous than *Nanog*.

To add quantification and clarity to this analysis, the co-occurrence of *Nanog* and *Oct4* signal where DAPI signal was present was plotted using the FIJI Coloc2 plug-in. This software also calculated the Pearson correlation coefficient (r^2), with higher r^2 values indicating increased correlation between *Nanog* and *Oct4* signals. These plots and correlation values mirrored the pattern that was implied from the immunofluorescence images. This is clearly demonstrated by the increase in r^2 value from 0.70 in serum E14 mESCs to 0.93 in 2i E14 mESCs, indicating a substantial increase in the co-occurrence of *Nanog* and *Oct4* signal following 2i adaptation (Figure 3.6 C, left panels). This reinforced the initial observation that *Nanog* expression becomes more homogeneous upon adaptation to 2i, meaning that there is an increased presence of both *Nanog* and *Oct4* in 2i cells compared to serum cells. Following reversion of 2i E14 mESCs into serum, the r^2 value returned to 0.72, reflecting the loss of homogeneous *Nanog* expression observed during reversion to serum culture.


A
E14
DAPI
Nanog (TRITC)
Oct4 (Far-red)
Merge
Serum
2i
Reversion


B
lsh-/-
DAPI
Nanog (TRITC)
Oct4 (Far-red)
Merge
Serum
2i
Reversion

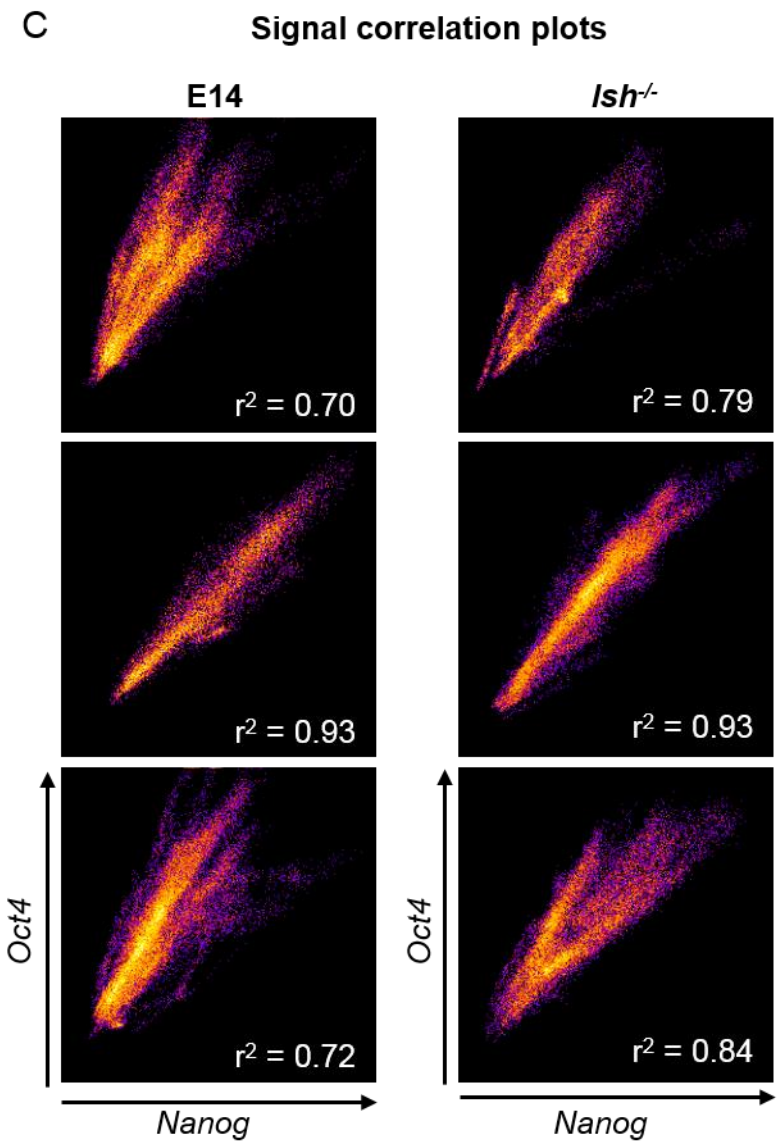


Figure 3.6. *Nanog* expression becomes more homogeneous upon adaptation to 2i and reverts to heterogeneous expression following reversion into serum in both E14 and *lsh*$^{-/-}$ mESCs. Immunofluorescence on E14 (A) and *lsh*$^{-/-}$ (B) mESCs cultured in serum, adapted to 2i for 15 days and reverted into serum culture for 15 days following 2i adaptation, using antibodies against *Nanog* and *Oct4*. Scale bars represent 50 µm. C. Scatter plots showing co-occurrence of *Nanog* and *Oct4* signal in the E14 and *lsh*$^{-/-}$ merge images. Generated using FIJI Coloc2 plug-in. Degree of signal correlation is indicated by r^2 value. The secondary antibody-only control staining relevant to this double immunofluorescence experiment is presented in Appendix Figure A1. C.

Double immunofluorescence to assess the uniformity of *Nanog* expression compared to *Oct4* was also performed on serum, 2i and reversion *lsh*$^{-/-}$ mESCs. The expression patterns of *Nanog* and *Oct4* in these cells emulated that of serum, 2i and reversion E14 mESCs. *Nanog* expression increased in uniformity following conversion from serum

to 2i, then returned to heterogeneous expression after reversion into serum, while *Oct4* expression was fairly uniform throughout (Figure 3.6 B). The signal correlation graphs plotting the co-occurrence of *Nanog* and *Oct4* signal reinforce these results. The r^2 value increased from 0.79 in serum *lsh*$^{-/-}$ mESCs to 0.93 after 2i adaptation, then decreased to 0.84 following reversion into serum (Figure 3.6 C, right panels). This further reflects the trend observed in E14 mESCs. It is notable that the r^2 values of serum and reverted *lsh*$^{-/-}$ mESCs are slightly higher than those of E14 mESCs, adding strength to the observation in Figure 3.5 A that *Nanog* expression is already more homogeneous in serum *lsh*$^{-/-}$ mESCs compared to E14 mESCs.

These analyses show that both E14 and *lsh*$^{-/-}$ mESCs exhibit increased homogeneity of *Nanog* and *Esrrb* expression upon adaptation to 2i, which is reversed once the cells are returned to serum culture. This strengthens the suggestion from the morphological analysis that both E14 and *lsh*$^{-/-}$ mESCs are able to successfully transition between serum and 2i culture. However, the information provided through analysis of *Nanog*, *Esrrb* and *Oct4* expression by immunofluorescence is limited and largely qualitative. Quantitative analysis of gene expression changes was needed to add extra weight to the observation that E14 and *lsh*$^{-/-}$ mESCs had adapted to and reverted from 2i culture successfully.

3.2.3.4 E14 and lsh$^{-/-}$ mESCs exhibit mRNA expression changes indicative of adaptation to and reversion from 2i

Studies into the molecular basis of ground state pluripotency demonstrate that adaptation to 2i is accompanied by expression changes in a number of developmentally-relevant genes (Marks *et al.*, 2012). The expression levels of a few key genes were monitored by quantitative reverse transcriptase-PCR (qRT-PCR) to quantitively assess whether E14 and *lsh*$^{-/-}$ mESCs had transitioned between serum and 2i on a transcriptional level. Representative qRT-PCR analyses of three technical replicates from one biological replicate are shown in Figure 3.7. Previous studies have indicated that the expression of pluripotency factors *Prdm14* and *Nanog* increase following conversion to 2i (Marks *et al.*, 2012; Wang *et al.*, 2014). This was found to be the case for both E14 and *lsh*$^{-/-}$ mESCs, with an approximate three-fold increase in *Nanog* expression and a five to eight-fold increase in *Prdm14* mRNA levels upon

adaptation to 2i (Figure 3.7 A). The mRNA levels of these pluripotency factors returned to levels observed prior to 2i adaptation in reverted E14 and *lsh*$^{-/-}$ mESCs. *Oct4* mRNA levels were also measured in serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs and were found not to change between culture conditions, reinforcing the *Oct4* staining shown in Figure 3.6 and expression data from previously published literature (Marks *et al.*, 2012). These results demonstrate that both E14 and *lsh*$^{-/-}$ mESCs undergo the expected expression changes in key pluripotency genes during adaptation to and reversion from 2i culture.

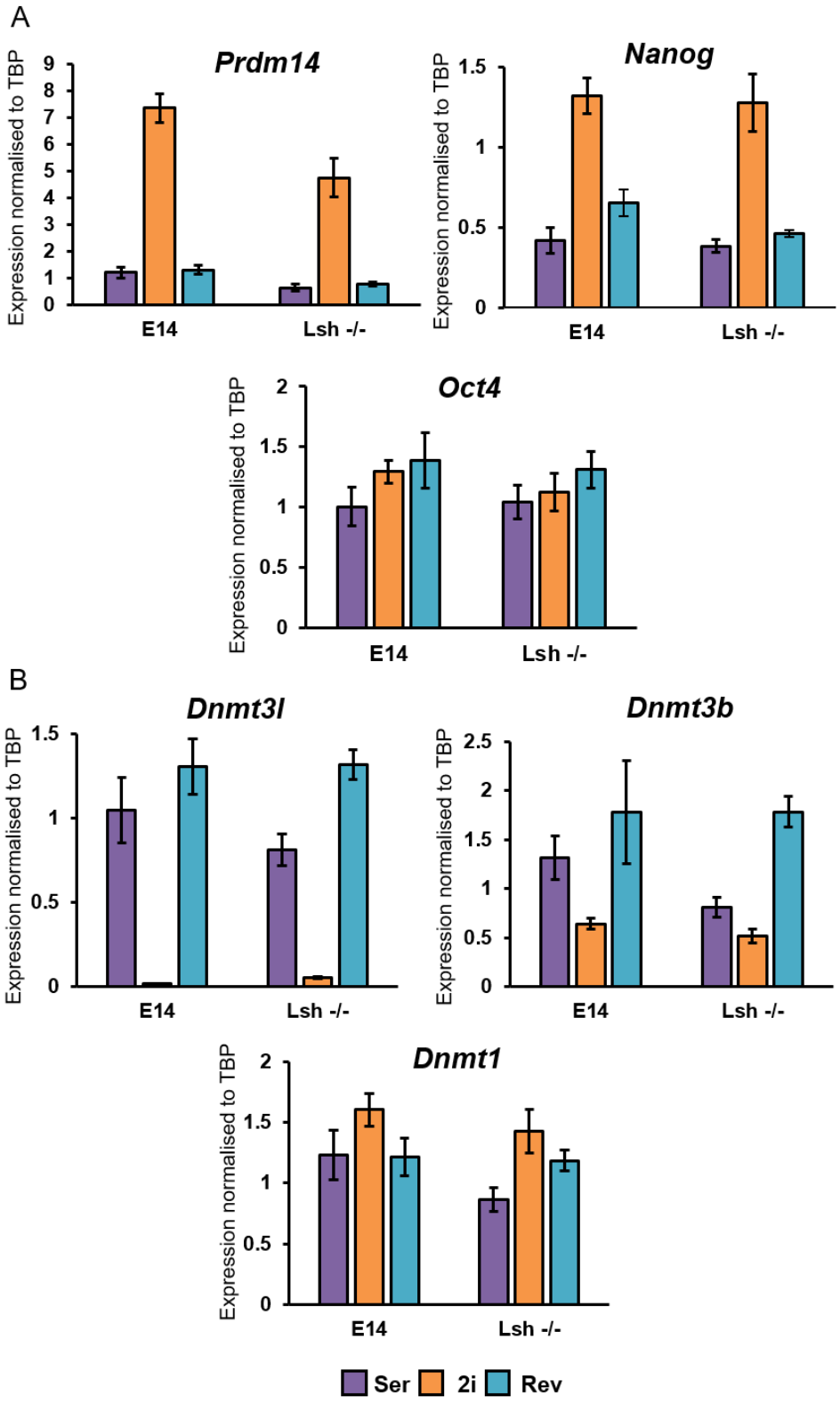
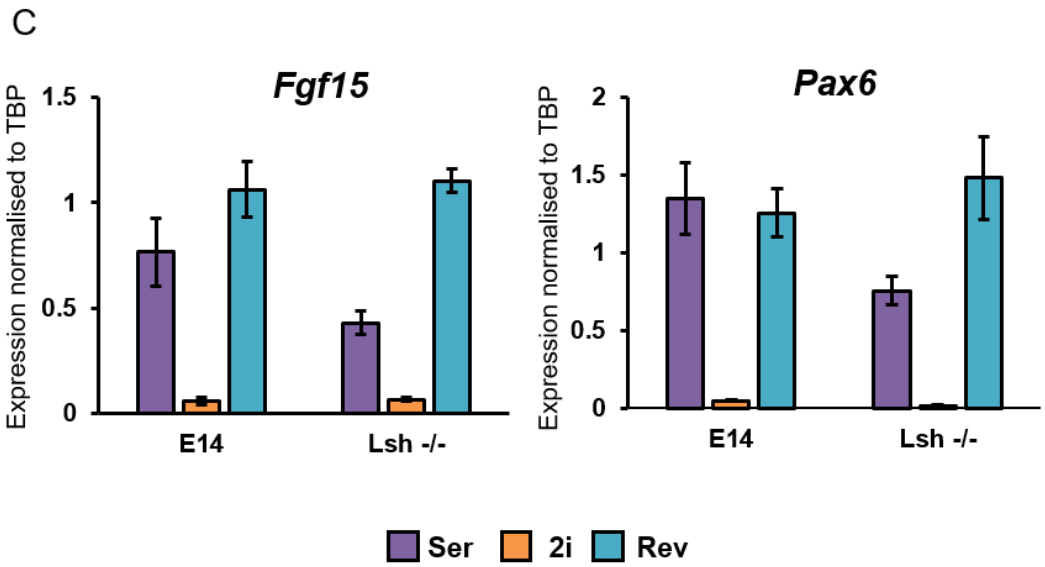


Figure 3.7. Key genes undergo expression changes in E14 and *Ish*$^{-/-}$ mESCs during adaptation to and reversion from 2i. qRT-PCR analysis of mRNA expression levels for key pluripotency factors (A), DNMTs (B) and lineage commitment-associated genes (C). Expression of markers normalised to TATA-binding protein (TBP) expression. Error bars represent +/- propagated standard deviation of 3 technical replicates.



This analysis was extended to encompass genes associated with DNA methylation deposition, which have been shown to markedly decrease in expression upon conversion to 2i (Marks *et al.*, 2012; Yamaji *et al.*, 2013). The expression levels of *de novo* DNA methyltransferase *Dnmt3b* and *de novo* methylation co-factor *Dnmt3l* were checked in serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs. *Dnmt3b* expression was reduced more than two-fold in E14 mESCs upon adaptation to 2i, and then recovered following reversion into serum (Figure 3.7 B). In serum *lsh*$^{-/-}$ mESCs prior to 2i adaptation, *Dnmt3b* levels are markedly lower than in serum E14 mESCs, meaning that the reduction in *Dnmt3b* expression after conversion to 2i is not as substantial. When these 2i *lsh*$^{-/-}$ mESCs are reverted into serum, *Dnmt3b* mRNA levels increase past those initially observed in serum and up to levels comparable to reverted E14 mESCs. *Dnmt3l* expression follows a similar pattern to *Dnmt3b* in both E14 and *lsh*$^{-/-}$ mESCs. *Dnmt3l* mRNA levels were lower in serum *lsh*$^{-/-}$ mESCs compared to serum E14 mESCs, but were still reduced to almost undetectable levels in both E14 and *lsh*$^{-/-}$ mESCs adapted to 2i. These mRNA levels then recovered following reversion into serum and, particularly in *lsh*$^{-/-}$ mESCs, exceeded the levels observed in serum prior to 2i adaptation. Expression levels of the maintenance methyltransferase *Dnmt1* were also measured in E14 and *lsh*$^{-/-}$ mESCs and were shown to remain relatively stable between serum, 2i and reversion culture conditions.

To further establish whether E14 and *lsh*$^{-/-}$ mESCs had adopted features of adaptation to 2i such as a reduction in lineage priming, the expression levels of early ectodermal

markers *Fgf15* and *Pax6* were assessed. Both markers followed a similar expression pattern in E14 and *lsh*$^{-/-}$ mESCs, with a substantial decrease in mRNA levels evident following 2i conversion, reflecting the reduced occurrence of lineage priming in this culture state (Figure 3.7 C). The expression levels of these lineage-associated genes then recovered after reversion into serum. Therefore, both E14 and *lsh*$^{-/-}$ mESCs exhibited expression changes in lineage markers associated with transitioning between serum and 2i culture states. However, the expression profiles of *Fgf15* and *Pax6* in *lsh*$^{-/-}$ mESCs differ in some aspects from E14 mESCs and mimic some previously observed features of *Dnmt3b* and *Dnmt3l* expression. For example, the mRNA levels of these genes are markedly lower in serum *lsh*$^{-/-}$ mESCs compared to E14 mESCs. Also, in reverted *lsh*$^{-/-}$ mESCs, expression levels of these genes increase beyond the level observed initially in serum, and up to levels comparable to E14 mESCs. This demonstrates that key genes involved in DNA methylation and lineage commitment are already downregulated in serum *lsh*$^{-/-}$ mESCs before conversion to 2i, implying that the expression profiles of serum *lsh*$^{-/-}$ mESCs are more similar to 2i than E14 mESCs. The downregulation of these genes in serum *lsh*$^{-/-}$ mESCs compared to E14 mESCs however seems to be lost during reversion into serum culture, suggesting that conversion to 2i acts to reset the transcriptional profile of *lsh*$^{-/-}$ mESCs. These observations are based on one biological replicate and need to be repeated for confirmation. Genome-wide expression analysis is also required to investigate whether *lsh*$^{-/-}$ mESCs resemble 2i mESCs transcriptionally on a global scale, and whether the genome-wide expression profile of *lsh*$^{-/-}$ mESCs is reprogrammed following reversion into serum.

Collectively, analysis of these key pluripotency, DNA methylation and lineage-associated markers shows that in terms of transcription, E14 and *lsh*$^{-/-}$ mESCs respond comparably during adaptation and reversion, despite there being differences in mRNA levels of certain genes prior to 2i conversion. Both mESC lines exhibit gene expression changes expected in cells reaching ground state pluripotency, therefore this quantitative analysis provides further evidence that E14 and *lsh*$^{-/-}$ mESCs are successfully adapting to and reverting from 2i culture.

3.2.3.5 Global DNA methylation levels fully recover in both E14 and *lsh*$^{-/-}$ mESCs following 2i-associated hypomethylation

Since I had established, using multiple experimental methods, that E14 and *lsh*$^{-/-}$ mESCs transition between serum and 2i culture morphologically and transcriptionally, I wanted to evaluate the impact of these transitions on DNA methylation levels. A few previous studies have described the global DNA methylation changes that occur during adaptation to and reversion from 2i. They demonstrate that conversion to 2i coincides with a two to three-fold reduction in global DNA methylation, which is subsequently fully recovered once cells are returned to serum culture (Ficz *et al.*, 2013; Habibi *et al.*, 2013; Leitch *et al.*, 2013). I used two different techniques to assess whether these DNA methylation changes occurred in the culture system I had developed. Initially, methylation levels at satellite DNA were assessed using the methylation-sensitive restriction endonuclease *MaeII* that digests unmethylated ACGT sites, which are conveniently enriched in major satellite regions of the genome (Abdurashitov *et al.*, 2009). Although *MaeII* restriction sites are enriched mainly in satellite DNA, digestion using this enzyme can provide a relatively quick indication of DNA methylation changes occurring globally due to the large proportion of the methylated genome that satellite regions comprise. Using this technique, *MaeII*-digested DNA is separated using agarose gel electrophoresis. DNA hypomethylation results in increased restriction digestion by *MaeII* and the subsequent appearance of lower molecular weight fragments on the agarose gel (indicated by the arrow in Figure 3.8 A).

Serum, 2i and reversion E14, WT (a wild-type clone from the CRISPR targeting) and *lsh*$^{-/-}$ mESCs were assayed using this method. Upon adaptation to 2i a number of low molecular weight bands appeared, accompanied by an overall downward shift in the molecular weight of the digested DNA, indicating de-methylation of satellite DNA (Figure 3.8 A). This de-methylation appears to be reversed following reversion into serum in E14, WT and *lsh*$^{-/-}$ mESCs, as demonstrated by the disappearance of low molecular weight fragments and the lack of downward shift in the molecular weight of the bulk DNA. This suggests that the absence of *Lsh* does not impede the re-establishment of DNA methylation, at least at these *MaeII* restriction sites. Restriction digestion with a methylation-insensitive restriction endonuclease, *MspI*, was also

performed and gel electrophoresed as a control for digestibility of the DNA, as well as uncut DNA to control for loading (Figure 3.8 B).

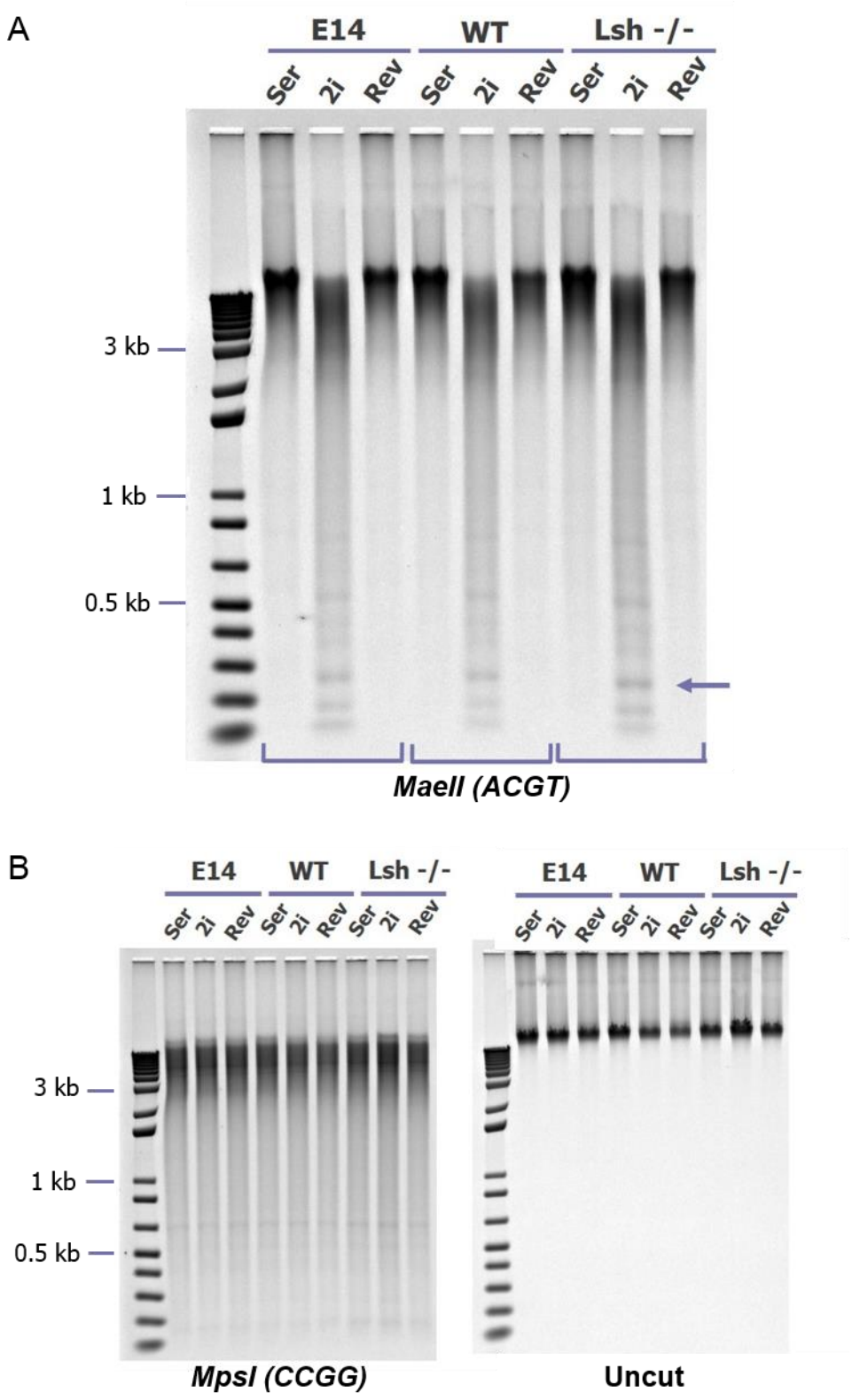


Figure 3.8. DNA methylation at major satellites is reduced upon 2i adaptation and re-established following serum reversion. A. Agarose gel displaying digestion of DNA from E14, WT (wild-type clone from the *Lsh* CRISPR targeting) and *lsh* $^{-/-}$ mESCs cultured in media containing serum (Ser), adapted to 2i for 15 days (2i) and returned into serum-containing media for 15 days (Rev) by methylation-sensitive *MaeII* restriction

enzyme. Arrow highlights presence of lower molecular weight bands, indicating hypomethylation of satellite DNA sequences. B. Agarose gels showing *MspI* digestions and uncut DNA for the same samples shown in B to validate genomic digestion compatibility, DNA integrity and equal loading.

Although this method can give an indication of changes in global DNA methylation, quantitative analysis is required to confirm the observations made from the methylation-sensitive digest. E14 and *lsh*$^{-/-}$ mESC DNA was digested to single nucleotides and 5-mC measured using liquid-chromatography mass spectrometry (LC-MS). LC-MS was performed by Jimi Wills in the Mass Spectrometry facility at the IGMM, with the analysis being carried out by myself with assistance from Jimi Wills and David Hay. LC-MS analysis of two biological replicates for each sample is shown in Figure 3.9. In E14 mESCs there was more than a two-fold reduction in global 5-mC following adaptation to 2i, which was subsequently fully recovered after reversion into serum (Figure 3.9). The same trend was observed for *lsh*$^{-/-}$ mESCs, as global DNA methylation fully recovered to levels present prior to 2i adaptation. The global DNA methylation levels of serum and reverted *lsh*$^{-/-}$ mESCs appeared to be lower than those of E14 mESCs, however these differences were not significant. This LC-MS analysis supports the suggestion from the methylation-sensitive DNA digest that the absence of *Lsh* does not affect the recovery of global DNA methylation levels during reversion from 2i culture.

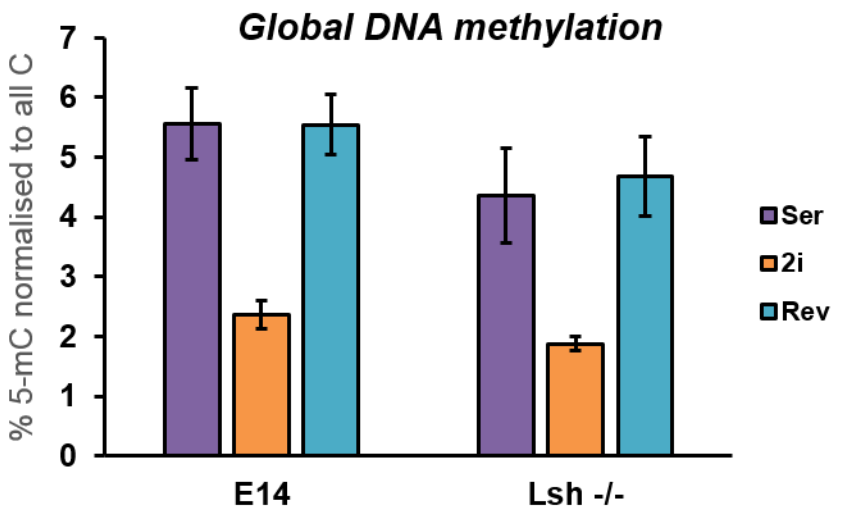


Figure 3.9. Global DNA methylation levels recover in both E14 and *Ish*$^{-/-}$ mESCs following 2i-associated hypomethylation. Quantitative analysis of global DNA methylation levels (5-mC) in serum (Ser), 2i and reversion (Rev) E14 and *Ish*$^{-/-}$ mESCs by LC-MS. Global 5-mC is calculated as a percentage of total cytosine levels (sum of LC-MS measurements for C, 5-mC and 5-hmC). Error bars represent +/- standard error of two biological replicates. All differences in global 5-mC levels between E14 and *Ish*$^{-/-}$ mESCs were found to be not significant by an unpaired two-way t-test.

Overall, these results demonstrate that DNA methylation levels are indeed globally reduced following adaptation to 2i, and that they can be recovered by returning the cells to serum culture. These analyses also suggest that the presence of *Lsh* is not necessary for this recovery of DNA methylation to occur. Therefore it appears, perhaps surprisingly, that *Lsh* does not participate in establishment of DNA methylation in mESCs during transition from 2i to serum culture.

3.2.4 Investigating the requirement for Dnmt3l in re-establishment of DNA methylation during reversion from 2i

My analysis suggests that *Lsh* does not appear to be required for re-methylation of the genome following the global hypomethylation associated with adaptation to 2i. I aimed to explore whether other factors may be involved in the re-establishment of DNA methylation during reversion from 2i. The function of *de novo* DNA methyltransferase co-factor *Dnmt3l* in mESCs is much less well characterised than that of its catalytically active counterparts. Therefore, I aimed to adapt and revert *dnmt3l*$^{-/-}$ mESCs over a 15-

day cycle to elucidate whether *Dnmt3l* is required for establishment of DNA methylation in this mESC culture system.

3.2.4.1 *WT and dnmt3l$^{-/-}$ mESCs morphologically and transcriptionally transition between serum and 2i*

To examine the effect absence of *Dnmt3l* has on global DNA methylation recovery following 2i-associated hypomethylation, a *dnmt3l$^{-/-}$* mESC line was kindly provided by the Oliviero laboratory. This mESC line was generated by targeting exon 1 of *Dnmt3l* using TALEN gene editing technology, as described in Neri *et al.* (2013). This generated an 11 bp deletion, creating a premature stop codon in exon 1. I confirmed this deletion using Sanger sequencing (Figure 3.10). A WT mESC line from the TALEN-targeting procedure was also provided, and I confirmed that these cells had the same *Dnmt3l* sequence as E14 mESCs. The absence of DNMT3L protein in *dnmt3l$^{-/-}$* mESCs was verified by Neri *et al*. (2013) using an antibody they had produced in-house. I endeavoured to confirm the loss of DNMT3L protein by Western blot, however attempts using commercially available antibodies were repeatedly unsuccessful (data not shown).

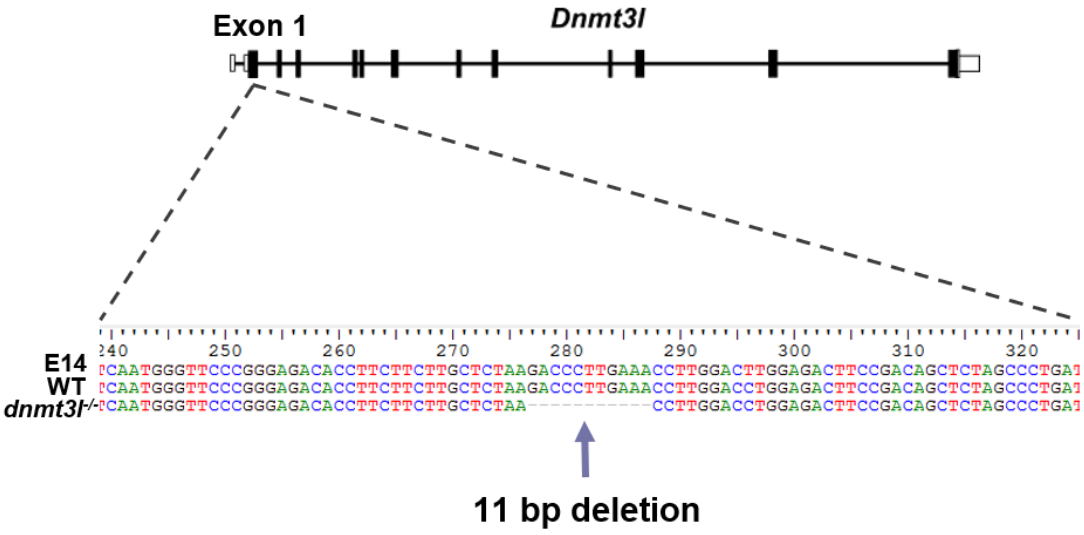


Figure 3.10. Confirmation of WT and *dnmt3l$^{-/-}$* mESCs. Exon 1 of the *Dnmt3l* gene was targeted for mutagenesis using TALEN gene editing technology, as described by Neri *et al.* (2013). Coding exons are shown in black. Alignment of sequences generated by Sanger sequencing of E14, WT and *dnmt3l$^{-/-}$* mESCs is shown, highlighting the 11 bp out-of-frame deletion present in *dnmt3l$^{-/-}$* mESCs.

WT and *dnmt3l*$^{-/-}$ mESCs were adapted to 2i and reverted into serum according to the experimental setup detailed in Section 3.2.3.1 (Figure 3.3). As with *lsh*$^{-/-}$ mESCs, changes in colony morphology were monitored to indicate adaptation to 2i. Both WT and *dnmt3l*$^{-/-}$ mESCs exhibited morphological changes reminiscent of adaptation to 2i (Figure 3.11). However, these mESC lines did not achieve typical 2i morphology to the same extent as E14 mESCs (shown in Figure 3.4), as although the colonies had become more three-dimensional with smoothened edges, the size and shape of the colonies did not become entirely homogeneous. Following reversion into serum culture, WT and *dnmt3l*$^{-/-}$ mESCs returned to a more flattened, heterogeneous appearance. Examination of colony morphology showed that both WT and *dnmt3l*$^{-/-}$ mESCs adopted some of the features indicative of transition between serum and 2i culture. However, the morphological changes observed in WT and *dnmt3l*$^{-/-}$ mESCs were not as obvious as those of E14 mESCs (shown in Figure 3.4).

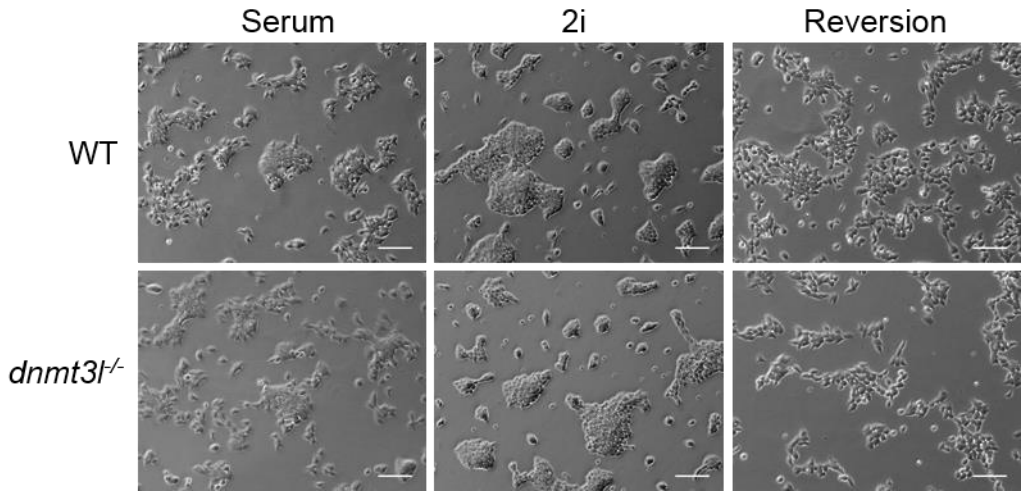


Figure 3.11. WT and *dnmt3l*$^{-/-}$ mESCs exhibit a degree of morphological transition between serum and 2i culture. Brightfield images of WT and *dnmt3l*$^{-/-}$ mESCs cultured in serum, adapted to 2i for 7 days (3 passages) and reverted into serum for 9 days (4 passages) following 15 days of 2i adaptation. Scale bars represent 100 µm.

To evaluate whether WT and *dnmt3l*$^{-/-}$ mESCs were adapting to and reverting from 2i in terms of transcription, the expression changes of a few key genes were analysed by qRT-PCR of three technical replicates from one biological replicate. Expression of pluripotency factor *Prdm14* increased in 2i WT and *dnmt3l*$^{-/-}$ mESCs as expected, then returned to levels similar to those observed initially in serum following reversion (Figure 3.12 A). *Dnmt3b* mRNA levels followed the expected trend in *dnmt3l*$^{-/-}$

mESCs, as they reduced substantially upon adaptation to 2i then increased back up to initial levels after reversion into serum (Figure 3.12 B). *Dnmt3b* expression also markedly reduced in 2i WT mESCs, however expression levels did not recover following reversion into serum, suggesting that WT mESCs did not transition between culture states as comprehensively as *dnmt3l*$^{-/-}$ mESCs. Expression levels of *Dnmt3l* itself were also checked in WT and *dnmt3l*$^{-/-}$ mESCs. *Dnmt3l* mRNA levels followed the expected pattern in WT mESCs, with a substantial reduction upon adaptation to 2i followed by a full recovery after reversion into serum (Figure 3.12 B). There appeared to be residual mRNA expression in *dnmt3l*$^{-/-}$ mESCs which also followed the same pattern, albeit on a much reduced scale. More extensive analysis is required to fully evaluate whether WT and *dnmt3l*$^{-/-}$ mESCs are transitioning between serum and 2i culture transcriptionally. However, despite a couple of irregularities, this expression data adds strength to the morphological analysis, suggesting that both mESC lines are successfully adapting to and reverting from 2i culture.

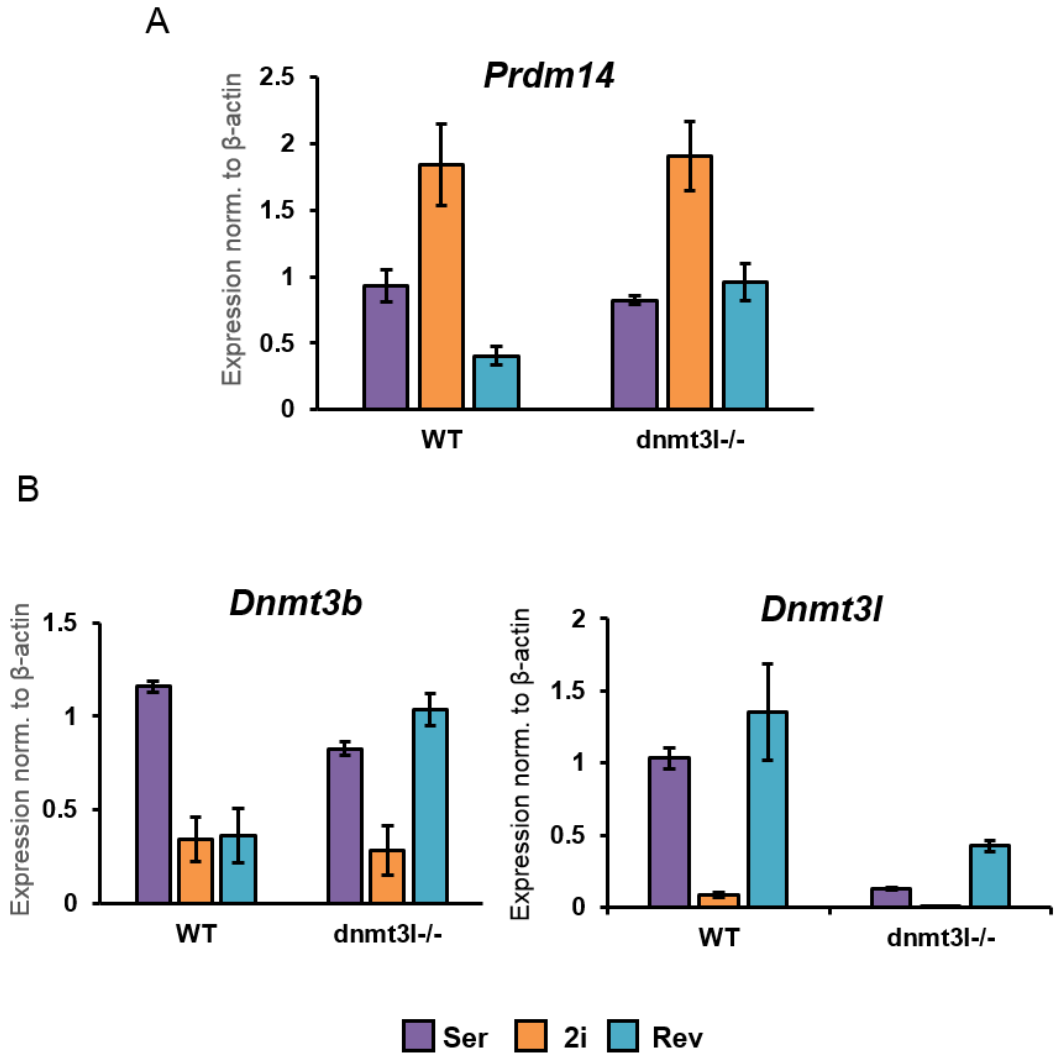


Figure 3.12. WT and *dnmt3l*$^{-/-}$ mESCs transition between serum and 2i culture transcriptionally. qRT-PCR analysis of mRNA expression levels of *Prdm14* (A), *Dnmt3b* and *Dnmt3l* (B) in serum (Ser), 2i and reversion (Rev) WT and *dnmt3l-/-* mESCs. Expression is normalised to *β-actin* mRNA levels. Error bars represent +/- propagated standard deviation of three technical replicates.

3.2.4.2 Re-establishment of DNA methylation following reversion from 2i is impeded in dnmt3l$^{-/-}$ mESCs

To investigate the effect of the absence of *Dnmt3l* on the recovery of DNA methylation following 2i-associated global hypomethylation, the levels of DNA methylation were examined in serum, 2i and reversion WT and *dnmt3l*$^{-/-}$ mESCs. Changes in DNA methylation levels were initially examined using a methylation-sensitive *MaeII* digest. Serum E14 mESCs were included as a comparison. WT mESCs in serum appeared to have similar DNA methylation levels to E14 mESCs (Figure 3.13 A). In contrast,

serum *dnmt3l*$^{-/-}$ mESCs exhibit noticeable hypomethylation, indicated by the downwards smear of the bulk DNA down the lane of the gel as well as the increased presence of lower molecular weight bands. Upon adaptation to 2i, there was no obvious change in the DNA methylation level of WT or *dnmt3l*$^{-/-}$ mESCs, at least at the satellite DNA regions that this technique assays. Moreover, there was no clear change in major satellite methylation following reversion of 2i WT and *dnmt3l*$^{-/-}$ mESCs into serum. This suggested that although WT and *dnmt3l*$^{-/-}$ mESCs apparently adapted to and reverted from 2i morphologically and transcriptionally, they did not transition at the level of DNA methylation at major satellite sequences.

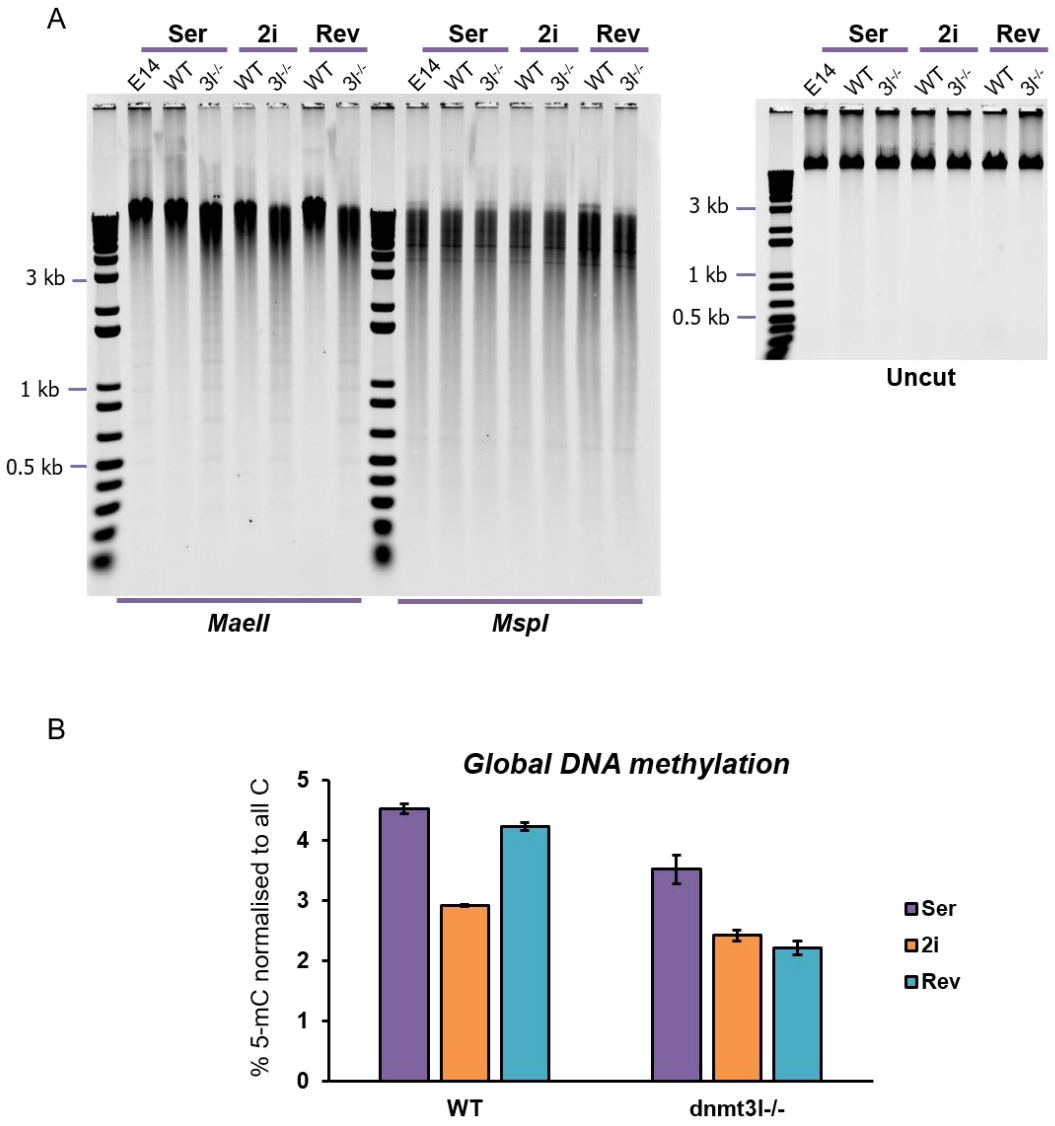


Figure 3.13. Global DNA methylation levels do not recover in *dnmt3l*$^{-/-}$ mESCs following 2i-associated hypomethylation. A. Agarose gel showing digestion of DNA from WT and *dnmt3l*$^{-/-}$ (*3l*$^{-/-}$) mESCs cultured in media containing serum (Ser), adapted to 2i for 15 days (2i) and returned into serum-containing media for 15 days (Rev) by methylation-sensitive *MaeII* restriction enzyme (left panel). Digestion of serum E14 mESCs is shown for comparison. Hypomethylation of satellite DNA sequences is indicated by smearing of the DNA down the lane of the gel and the appearance of lower molecular weight fragments. *MspI* digests are shown as a control for digestability (left panel). Uncut genomic DNA is also shown as a loading control (right panel). B. Quantitative analysis of global DNA methylation levels (5-mC) by LC-MS. Global 5-mC is calculated as a percentage of total cytosine levels (sum of LC-MS measurements for C, 5-mC and 5-hmC). Error bars represent +/- standard deviation of three technical replicates.

To corroborate this surprising result, LC-MS was performed on serum, 2i and reversion WT and *dnmt3l*$^{-/-}$ mESCs (three technical replicates from one biological

replicate for each sample) to measure global DNA methylation levels. This revealed that there was in fact a reduction of approximately 35% in global 5-mC levels in WT mESCs upon adaptation to 2i (Figure 3.13 B). This reduction was almost entirely recovered following reversion into serum. In *dnmt3l*$^{-/-}$ mESCs, global DNA methylation levels were already reduced by approximately 20% in serum compared to WT mESCs. Following conversion to 2i, global 5-mC levels decreased a further 30%. This loss of DNA methylation was not recovered after reversion of *dnmt3l*$^{-/-}$ mESCs back into serum.

These results demonstrate that the expected 2i-associated global hypomethylation did in fact occur in WT and *dnmt3l*$^{-/-}$ mESCs, albeit not to the same extent as in E14 mESCs. This DNA methylation reduction was not apparent in the methylation-sensitive DNA digests, possibly due to the satellite DNA sequences assessed by this method remaining methylated in serum and 2i mESCs. Importantly, these results also highlight the complete inability of *dnmt3l*$^{-/-}$ mESCs to re-methylate the genome during reversion from 2i culture. This reveals, as predicted by published work (Grabole *et al.*, 2013; Leitch *et al.*, 2013; Yamaji *et al.*, 2013b), that *Dnmt3l* is an essential factor for re-establishing global DNA methylation levels in this mESC culture system. More in-depth analysis is required to verify this interesting result and to explore what regions of the genome remain unmethylated in reverted *dnmt3l*$^{-/-}$ mESCs. Further investigation into this result may also help to uncover the exact mechanism by which DNA methylation is re-established during transition from 2i to serum culture.

3.3 Discussion

The work presented in this chapter was undertaken with two main aims in mind: to develop a culture system in which to modulate global DNA methylation and to determine whether *Lsh* is required for DNA methylation re-establishment in this system. To achieve these aims, I optimised the transition of E14 and *lsh*$^{-/-}$ mESCs between serum and 2i culture to modulate DNA methylation levels. This optimisation involved monitoring colony morphology, as well as changes in expression levels and uniformity of expression of a few candidate genes to assess whether the mESCs were transitioning between culture states. DNA methylation levels were then measured at satellite regions and globally to examine the genome-wide hypomethylation and

subsequent re-methylation. This allowed me to assess the involvement of *Lsh* in this methylation process by transitioning E14 and *lsh*$^{-/-}$ mESCs through this culture set-up. Furthermore, once this culture system had been optimised I was able to adopt it to investigate the requirement for other factors during the re-methylation process, such *de novo* methyltransferase co-factor *Dnmt3l*.

3.3.1 Colony morphology and gene expression changes are good indicators of adaptation to and reversion from 2i

During development of this culture system, I needed a way to assess whether the transition between the culture states, and therefore changes in global DNA methylation, had occurred. Changes in colony morphology were monitored as a quick and easy way to give an initial indication as to whether the mESCs were responding to the change in culture conditions. Both E14 and *lsh*$^{-/-}$ mESCs exhibited clear changes in colony morphology after adaptation to 2i and reversion into serum, albeit *lsh*$^{-/-}$ mESCs to a slightly lesser extent. This corresponded to the anticipated changes in global DNA methylation levels, as shown in Section 3.2.3.5. The morphology of WT and *dnmt3l*$^{-/-}$ mESCs were also assessed and were shown not to change to the same degree as E14 or *lsh*$^{-/-}$ mESCs. Likewise, DNA methylation levels in WT and *dnmt3l*$^{-/-}$ mESCs were shown not to change to the same extent as in E14 and *lsh*$^{-/-}$ mESCs. Therefore, monitoring morphological changes could provide a straightforward and effective way to gauge the degree of adaptation of mESCs to the different culture states and perhaps predict whether the anticipated DNA methylation changes will occur. However, caution is required when drawing conclusions from morphological information, as the appearance of different mESC lines varies, which will affect their morphological adaptation during transitions between culture states.

Expression analysis of a few key pluripotency, lineage-associated and DNMT genes was also used as an effective indicator of 2i adaptation and serum reversion in E14 and *lsh*$^{-/-}$ mESCs. Care is also required when rationalising expression data, as most of the anticipated expression changes took place in WT and *dnmt3l*$^{-/-}$ mESCs after conversion to 2i. However, this was not accompanied by the expected two to three-fold reduction in DNA methylation levels, as the global methylation change was much smaller in magnitude than anticipated.

Overall, through this set of experiments I showed that monitoring colony morphology and gene expression changes provides a quick and effective way to assess the degree of adaptation to and reversion from 2i for a particular mESC line. However, caution is necessary when interpreting the data from these approaches as they may not be entirely accurate in predicting whether genome-wide DNA methylation changes have occurred.

3.3.2 Differences in colony morphology, gene expression and global DNA methylation levels are interconvertible between serum and 2i culture

Previously published studies suggested that serum and 2i culture states are interconvertible in terms of colony morphology, gene expression differences and crucially, DNA methylation profiles (Marks *et al.*, 2012; Habibi *et al.*, 2013). This feature of the culture system was key to enable the investigation into the requirement for *Lsh* in DNA methylation re-establishment in a pluripotent context. During the optimisation of this culture system, I demonstrated that these two culture states indeed appeared to be interconvertible in the context of colony morphology, gene expression and global DNA methylation levels. Changes in DNA methylation levels followed the anticipated pattern of genome-wide de-methylation upon adaptation to 2i. This global hypomethylation was then reversed after returning mESCs to serum culture, highlighting the interconvertibility of DNA methylation in this culture system. This also confirmed that it was possible to modify global DNA methylation levels by altering culture conditions. However, DNA methylation changes were only examined on a global scale or at major satellite sequences. It would be interesting to scrutinize the differences in genome-wide DNA methylation profiles between serum, 2i and reverted mESCs to evaluate whether DNA methylation is re-established to the same loci during transition from 2i to serum. Similarly, it would be interesting to determine whether gene expression profiles of reverted mESCs were entirely comparable to their serum counterparts, as is suggested by previous reports (Marks *et al.*, 2012). This would help to uncover whether the serum and 2i culture states are truly interconvertible.

3.3.3 Serum *lsh*$^{-/-}$ mESCs display some characteristics similar to 2i E14 mESCs

Whilst optimising the adaptation and reversion culture system, it became clear that serum *lsh*$^{-/-}$ mESCs showed some similarities to 2i E14 mESCs in terms of gene expression. This was most evident when examining the expression of lineage-associated factors *Pax6* and *Fgf15* as well as *Dnmt3b* and *Dnmt3l* (Figure 3.7 B and C). For each of these genes, the expression levels in serum *lsh*$^{-/-}$ mESCs were reduced approximately two-fold in comparison to serum E14 mESCs, representing an intermediate expression level between serum and 2i E14 mESCs. Additionally, analysis of uniformity of *Nanog* expression by immunofluorescence suggested that *Nanog* expression was already fairly homogeneous prior to 2i adaptation, meaning that the expected increase in uniformity was not as apparent in *lsh*$^{-/-}$ mESCs compared to E14 mESCs following conversion to 2i (Figures 3.5 A and 3.6). These observations imply that transcriptionally, *lsh*$^{-/-}$ mESCs could represent an intermediate state between serum and 2i, although more in-depth and genome-wide expression analysis is required to support these observations. To address this, RNA-seq was performed on serum and 2i E14 and *lsh*$^{-/-}$ mESCs and is discussed in Chapter 4. Furthermore, it would be interesting to investigate whether these similarities between serum *lsh*$^{-/-}$ mESCs and 2i E14 mESCs extended to the DNA methylation profiles of these mESCs. This was hinted to in the global DNA methylation analysis by LC-MS, as the global DNA methylation levels appeared to be lower in serum *lsh*$^{-/-}$ mESCs compared to serum E14 mESCs, however this difference was not found to be significant (Figure 3.9). Higher-resolution scrutinization of genome-wide DNA methylation profiles is necessary to further investigate this interesting observation.

3.3.4 Global DNA methylation levels fully recover after reversion into serum in *lsh*$^{-/-}$ but not in *dnmt3l*$^{-/-}$ mESCs

The primary aim of this chapter was to investigate the impact of absence of *Lsh* on the ability to re-establish DNA methylation following 2i-associated global hypomethylation. This was achieved by harnessing the interconvertibility of the 2i adaptation and serum reversion culture system. Using this culture system, I discovered that surprisingly, *Lsh* is not required to re-establish global levels of DNA methylation

during reversion from 2i to serum. Using *dnmt3l*$^{-/-}$ mESCs I then went onto show that *Dnmt3l* does appear to be required for re-methylation of the genome following 2i-associated global hypomethylation. This strengthens the finding that *Lsh* is not required for re-establishment of DNA methylation during transition to serum by providing an example of a factor that is required. In this way, *dnmt3l*$^{-/-}$ mESCs can act as a 'positive control', demonstrating that it is possible for global DNA methylation levels not to undergo full recovery following reversion into serum.

These findings have implications for the roles of these factors in DNA methylation deposition during the early stages of embryonic development. mESCs adapted to 2i are proposed to closely resemble cells of the ICM in terms of DNA methylation profiles (Habibi *et al.*, 2013; Choi *et al.*, 2017). In contrast, serum mESCs are proposed to reflect a slightly later, more 'primed' stage of development in terms of gene expression and DNA methylation levels. Therefore, the establishment of DNA methylation that occurs during transition from 2i to serum culture could be said to be representative of the *de novo* DNA methylation that occurs during the early stages of implantation of the blastocyst *in vivo* (Habibi *et al.*, 2013; Choi *et al.*, 2017). This implies that *Lsh* function is not required for DNA methylation deposition during this early stage of embryonic development. This finding provides new evidence to support a role for *Lsh* in establishing DNA methylation in a later developmental context. It also shows that although *Lsh* has been shown to be expressed more highly in pluripotent cultures, this does not translate into a role for *Lsh* in DNA methylation in these systems. In contrast, these experiments suggest that *Dnmt3l* is necessary for DNA methylation establishment in this culture model of early embryonic development, as had been previously suggested (Grabole *et al.*, 2013; Leitch *et al.*, 2013; Yamaji *et al.*, 2013) highlighting the role for this DNA methyltransferase co-factor in *de novo* methylation during this early phase of development.

Overall, the results discussed in this chapter provide valuable insight into the factors required for establishment of DNA methylation in mESCs. Further investigations are required to determine the stage of development where *Lsh* function begins to contribute to DNA methylation establishment. Moreover, it would be interesting to examine the requirement for *Dnmt3l* in re-methylation of the genome during reversion

in more detail as doing so could help to elucidate the mechanism by which DNA methylation is re-established in this culture model, as well as the factors involved in this process.

Chapter 4. Mapping genome-wide DNA methylation and gene expression changes during transition to and from the pluripotent ground state

4.1 Introduction

A number of previous studies have demonstrated the interconvertibility between the transcriptional and DNA methylation profiles of mESCs cultured in conditions containing serum or 2i (Marks *et al.*, 2012; Habibi *et al.*, 2013; Leitch *et al.*, 2013). The ability to simply and effectively transition mESCs between these distinct pluripotent states by changing the culture conditions has provided a model in which to study the transcriptional and epigenetic mechanisms underlying the transitions between pluripotent states in early development. In particular, the extensive remodelling of genome-wide DNA methylation that occurs during transition between serum and 2i culture states is a valuable feature of this culture system, allowing dissection of the mechanisms involved in global de-methylation and subsequent DNA methylation re-establishment in a pluripotent context.

In Chapter 3, I harnessed the interconvertibility of this culture system to examine the impact of absence of *Lsh* on DNA methylation re-establishment during transition from the hypomethylated 2i state to hypermethylated serum mESCs. This revealed that *Lsh* does not appear to be required for global re-establishment of DNA methylation during this transition. Indeed, I showed that lack of *Lsh* appeared to have little effect on the apparent interconvertibility between the serum and 2i culture states in terms of colony morphology, transcription of a few key pluripotency and lineage-associated factors, and global DNA methylation levels. However, this analysis was limited to assessing the transcription of a few genes and bulk 5-mC levels, which are not informative of changes in genome-wide patterns of transcription and DNA methylation. Therefore, I aimed to inspect the genome-wide changes in gene expression and DNA methylation of *lsh*$^{-/-}$ mESCs during transition between serum and 2i culture states to uncover any previously unclear roles for *Lsh* in influencing patterns of gene expression or DNA methylation. Additionally, I sought to examine transcriptional and DNA methylation

profiles of serum, 2i and reversion E14 mESCs to further investigate whether gene expression and 5-mC patterns are truly interconvertible between the serum and 2i culture. To enable in-depth analysis of changes in transcriptomes and DNA methylomes between culture states, high-resolution transcription and 5-mC profiles were generated for serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs using RNA-seq and enhanced RRBS (ERRBS) technologies.

4.2 Results

4.2.1 Transcriptome analysis by RNA-seq

RNA-seq was employed to assess the transcriptional changes that occur in E14 and *lsh*$^{-/-}$ mESCs during adaptation to and reversion from 2i culture. I isolated polyadenylated mRNA and prepared libraries from one biological replicate for each cell line/condition (six samples in total) for Illumina sequencing, which was performed by Edinburgh Genomics. All bioinformatic analysis was performed by Dr Donncha Dunican.

4.2.1.1 Gene expression profiles of E14 and lsh$^{-/-}$ mESCs cluster based on culture conditions rather than genotype

Initially, the relationship between the global transcriptomes of E14 and *lsh*$^{-/-}$ serum, 2i and reversion mESCs was assessed. This revealed that gene expression profiles of these mESCs cluster based on the conditions in which they are cultured, rather than the presence or absence of *Lsh*. As illustrated by hierarchical cluster analysis (HCA), the transcriptomes of E14 and *lsh*$^{-/-}$ mESCs initially branch off depending on whether they were cultured in conditions containing serum (including serum and reversion mESCs) or 2i (Figure 4.1 A). The transcriptomes of E14 and *lsh*$^{-/-}$ serum and reversion mESCs then differ based on whether they were always cultured in serum, or were adapted to 2i then reverted into serum. This analysis demonstrates that mESCs cultured in 2i occupy a transcriptional profile distinct from those cultured in serum. Moreover, it shows that when mESCs are reverted into serum culture following 2i adaptation, their transcriptomes return to profiles that resemble, but are not identical to, those observed prior to 2i conversion. Importantly, this clustering analysis emphasises that

within each culture condition, the gene expression profiles of E14 and *lsh*$^{-/-}$ mESCs are closely related.

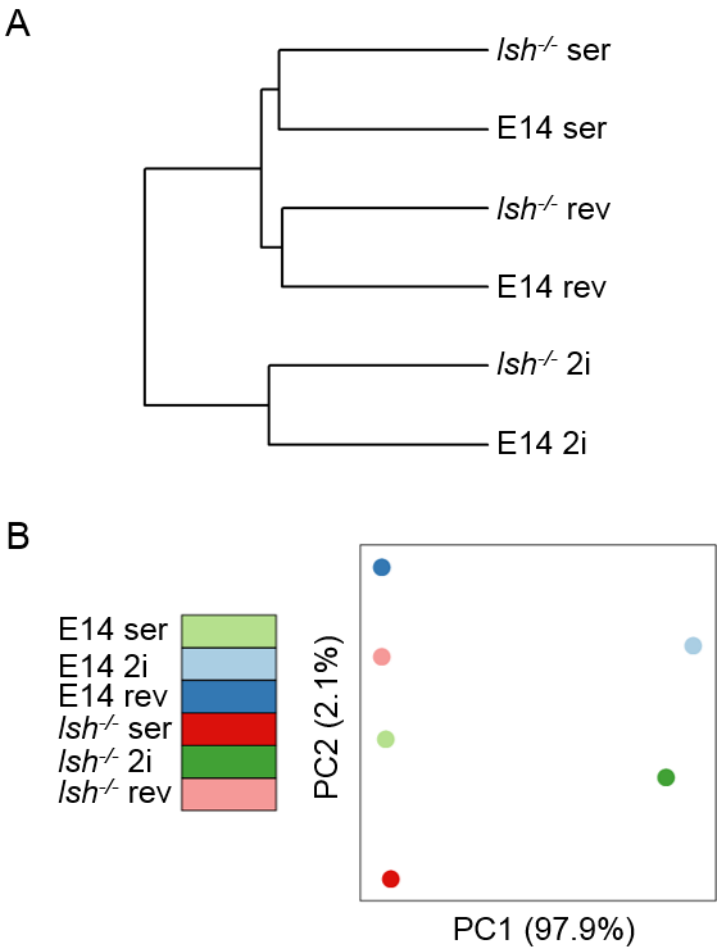


Figure 4.1. Gene expression profiles of E14 and *lsh*$^{-/-}$ mESCs cluster based on culture conditions rather than genotype. Analysis of the degree of similarity between serum, 2i and reversion E14 and *lsh*$^{-/-}$ RNA-seq datasets using hierarchical clustering analysis (A) and principal component analysis (B).

These initial indications were reinforced using principal component analysis (PCA), which further demonstrated that the transcriptional profiles of E14 and *lsh*$^{-/-}$ mESCs are comparable when grown under the same conditions (Figure 4.1 B). As implied by the HCA, the vast majority (97.9%) of the variance between the gene expression profiles analysed can be accounted for by transcriptional differences between mESCs cultured in 2i to those cultured in serum. The differences in gene expression between serum and reversion mESCs explains only 2.1% of the variance observed, further illustrating that these expression profiles closely resemble one another.

Overall, these analyses clearly demonstrate that the adaptation of E14 and *lsh*$^{-/-}$ mESCs to 2i culture induces extensive transcriptional changes, resulting in 2i mESCs that adopt an entirely distinct transcriptional state compared to the serum mESCs from which they were derived. It also highlights that this global adaptation of gene expression patterns is dynamic, as when 2i mESCs are reverted into serum culture, their transcriptional profiles also transition back to those resembling serum mESCs prior to 2i adaptation. Crucially, the absence of *Lsh* appears to have no observable effect on these gross transcriptional changes that occur upon adaptation to 2i and reversion back into serum, as the transcriptomes of *lsh*$^{-/-}$ mESCs closely resemble those of E14 mESCs in each culture condition.

4.2.1.2 Gene expression profiles of E14 and lsh$^{-/-}$ mESCs cultured in serum and 2i are generally interconvertible

To further characterise the global transcriptional changes that occur upon adaptation to 2i, the expression profiles of serum and 2i E14 mESCs were compared to one another in a scatter plot. This supported the indication from the previous analysis that mESCs undergo extensive genome-wide transcriptional changes upon adaptation to 2i, as more than 2700 genes were upregulated and more than 3000 downregulated at least two-fold in E14 2i mESCs compared to E14 serum (Figure 4.2 A). When serum E14 mESCs were plotted against reversion E14 mESCs, far fewer genes were differentially expressed between the two conditions, reinforcing the previous observation that gene expression profiles of 2i mESCs largely return to those of serum mESCs after reversion from 2i culture. This indicates, along with the PCA, that the transcriptomes of serum and 2i mESCs are largely interconvertible, dynamically transitioning between two states depending on culture conditions. However, there remain a considerable number of genes that are mis-expressed in reversion E14 mESCs compared to serum E14 mESCs (1262 upregulated genes and 994 downregulated genes), suggesting the two states are not entirely interconvertible in terms of transcription.

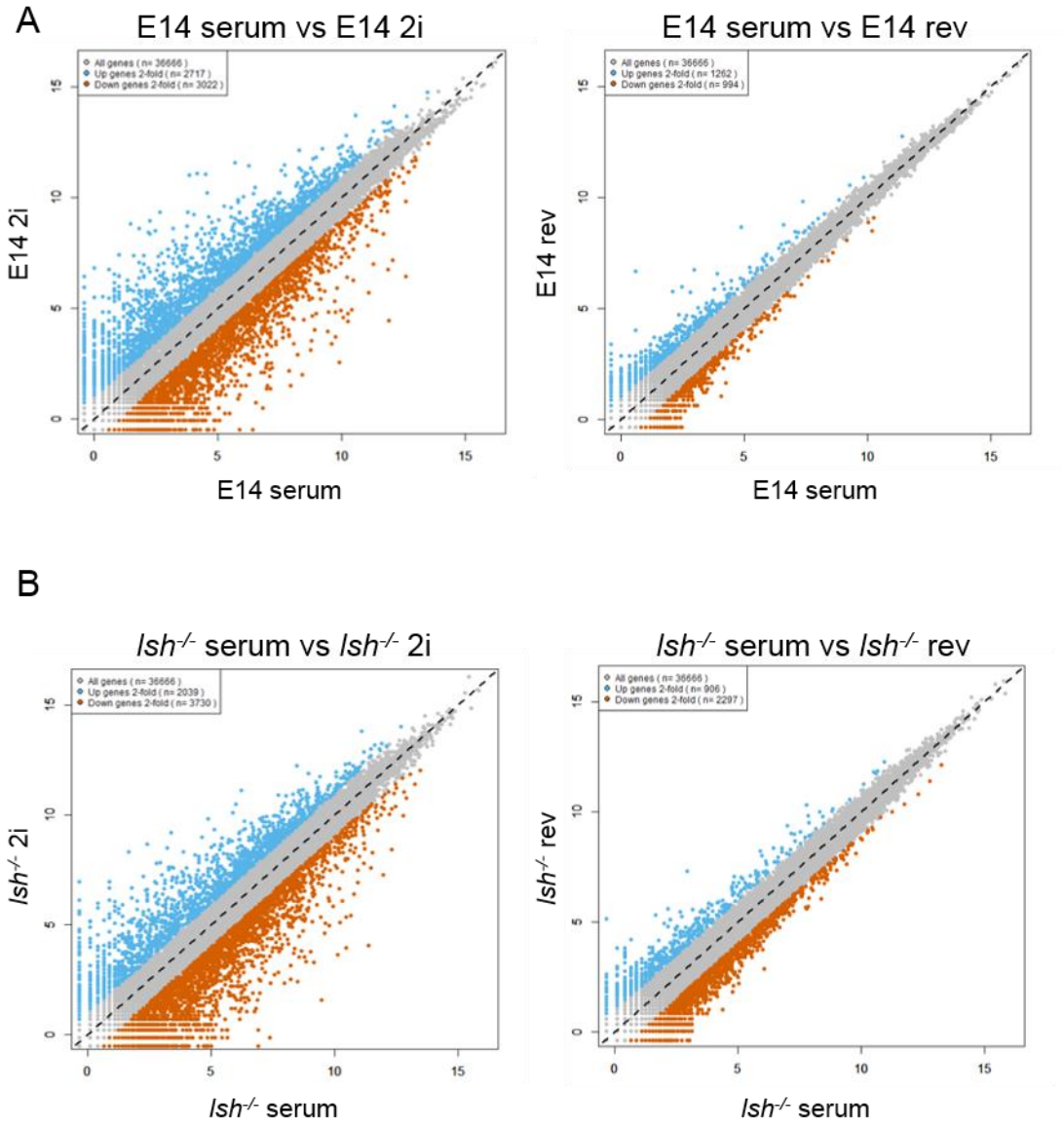


Figure 4.2. Gene expression profiles of E14 and *lsh*$^{-/-}$ mESCs cultured in serum or 2i are generally interconvertible. Scatter plots comparing RNA-seq profiles of serum, 2i and reversion (rev) E14 (A) and *lsh*$^{-/-}$ (B) mESCs. Genes upregulated more than two-fold are highlighted in blue, genes downregulated more than two-fold are shown in orange. Axes represent $\log_2$(read counts per 10 million) for the sample indicated on each axis.

A similar trend in the transcriptional changes between serum and 2i culture is also apparent in *lsh*$^{-/-}$ mESCs. As observed in the equivalent comparison in E14 mESCs, a large number of genes are upregulated and downregulated more than two-fold when serum *lsh*$^{-/-}$ mESCs are compared to those adapted to 2i (2039 upregulated, 3730 downregulated; Figure 4.2 B). Furthermore, there are far fewer differentially expressed genes when serum *lsh*$^{-/-}$ mESCs are compared to reverted *lsh*$^{-/-}$ mESCs (906

upregulated, 2297 downregulated). However, there are substantially more transcriptional differences between serum *lsh*$^{-/-}$ mESCs and reversion *lsh*$^{-/-}$ mESCs compared to the equivalent comparison in E14 mESCs, particularly evident in the number of downregulated genes in reverted mESCs compared to serum (2297 downregulated genes in reversion *lsh*$^{-/-}$ mESCs compared to 994 in reversion E14 mESCs). This suggests that although *lsh*$^{-/-}$ mESCs exhibit a degree of transcriptome interconvertibility between serum and 2i culture, the absence of *Lsh* somewhat hampers the upregulation of many genes during reversion into serum culture. Therefore, the extent of the transcriptome interconversion may not be as pronounced in *lsh*$^{-/-}$ mESCs as in E14 mESCs, but both mESC lines do exhibit a degree of interconvertibility in gene expression profiles between serum and 2i culture.

4.2.1.3 Serum lsh$^{-/-}$ mESCs show similar expression of a subset of genes to E14 2i mESCs

In Chapter 3, qRT-PCR and immunofluorescence analysis revealed that the expression of a few key pluripotency, lineage-associated and DNMT genes differed between serum E14 and *lsh*$^{-/-}$ mESCs, and that the expression of these factors in serum *lsh*$^{-/-}$ mESCs were closer to the levels observed in 2i E14 mESCs. To investigate whether this similarity in gene expression patterns between E14 2i and *lsh*$^{-/-}$ serum mESCs is applicable genome-wide, the transcriptional profiles of these mESCs obtained using RNA-seq were scrutinised.

The genes that changed in expression more than two-fold in serum *lsh*$^{-/-}$ mESCs compared to serum E14 mESCs were defined (775 genes). A heatmap was then compiled displaying the expression of this refined set of genes across serum and 2i E14 mESCs and serum *lsh*$^{-/-}$ mESCs (Figure 4.3 A). This revealed that, based on this refined dataset, the expression profiles of E14 2i and *lsh*$^{-/-}$ serum mESCs cluster together. This contrasts with the clustering analysis of the whole transcriptomes shown in Figure 4.1, where E14 and *lsh*$^{-/-}$ serum mESCs cluster together. This suggests that the genes that are mis-expressed in serum *lsh*$^{-/-}$ mESCs compared to serum E14 mESCs are the same genes that change expression in E14 mESCs during transition to 2i, reinforcing the suggestion from the qRT-PCR data in Chapter 3 that a subset of genes exhibit similar expression between serum *lsh*$^{-/-}$ and 2i E14 mESCs.

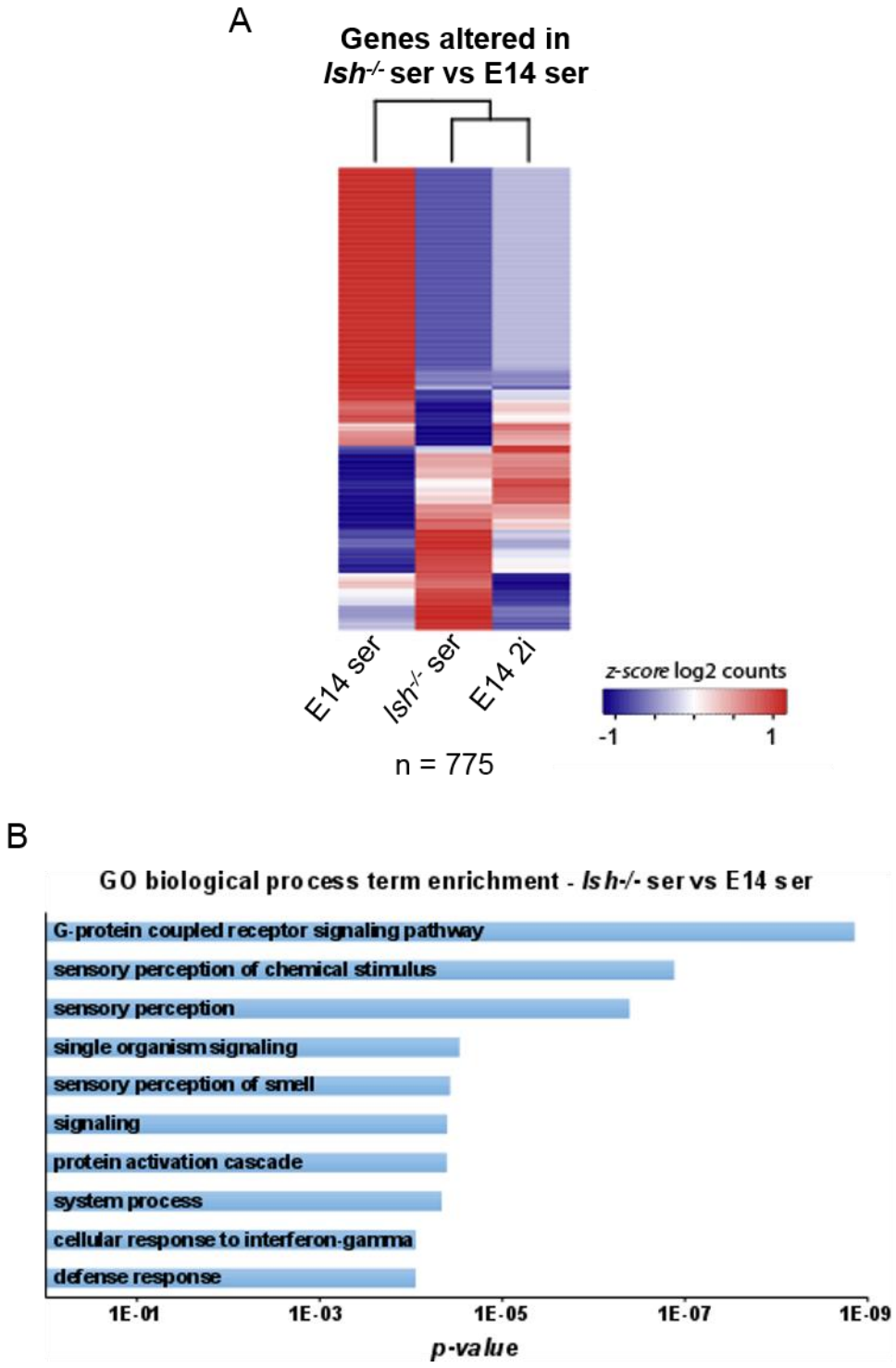


Figure 4.3. Serum *lsh*$^{-/-}$ mESCs show similar expression of a subset of genes to E14 2i mESCs. A. Heatmap analysis showing the expression levels of genes mis-expressed more than two-fold in serum *lsh*$^{-/-}$ mESCs compared to serum E14 mESCs. The expression levels of this defined gene set (775 genes) were then compared across serum and 2i E14 mESCs and serum *lsh*$^{-/-}$ mESCs. B. Gene ontology (GO) term enrichment analysis using the biological processes category with the GOrilla tool. The top ten significant terms are shown.

To determine whether the set of genes displayed in the heatmap are associated with any specific classes of genes, they were subjected to gene ontology (GO) enrichment analysis using the biological processes category of the GOrilla tool (http://cbl-gorilla.cs.technion.ac.il/). This revealed a significant enrichment for GO terms associated with signalling processes and sensory perception (Figure 4.3 B, top ten significantly enriched terms are shown). This demonstrates that genes mis-expressed in serum *lsh*$^{-/-}$ mESCs are involved in signalling and sensory perception processes, suggesting that *Lsh* may play a role in influencing expression of these classes of genes.

4.2.2 DNA methylome analysis by ERRBS

To examine the changes in DNA methylation that occur during transition to and reversion from 2i in E14 and *lsh*$^{-/-}$ mESCs, high-resolution 5-mC profiles were generated using ERRBS. RRBS was originally developed to provide single-nucleotide resolution DNA methylation profiles of CG-rich genomic regions (Meissner *et al.*, 2005). This enrichment of CG-rich sequences enabled coverage of the majority of promoters and other genomic features of interest in the genome, while reducing the overall amount of sequencing required compared to techniques such as whole genome bisulfite sequencing (WGBS). The generation of CG-rich libraries for RRBS involves digestion of genomic DNA with a methylation-insensitive enzyme, such as *MspI*, followed by size selection to purify CG-rich DNA fragments (Meissner *et al.*, 2005; Gu *et al.*, 2011). These DNA fragments are then subjected to bisulfite treatment which distinguishes unmodified cytosine from 5-mC and 5-hmC, as the treatment deaminates unmodified cytosine to uracil, while methylated and hydroxymethylated cytosines are protected from deamination. Following the sequencing of bisulfite treated DNA, unmethylated cytosines can be identified as thymine residues, while methylated cytosines are recognised as cytosine residues. Combination of these techniques with next-generation sequencing strategies resulted in the generation of high-quality single-nucleotide resolution DNA methylation profiles for a lower cost and input requirements compared to whole-genome based strategies (Meissner *et al.*, 2005; Gu *et al.*, 2011). However, these advantages come with the caveat of reduced genomic coverage that is largely restricted to CG-rich sequences. A modified RRBS protocol, where the size range of DNA fragments selected was increased, doubled the genomic

coverage of the technique and allowed examination of regions beyond CpG islands and promoters, such as exons, introns and other genomic regions (Akalin *et al.*, 2012; Garrett-Bakelman *et al.*, 2015). This improved method, termed enhanced RRBS (ERRBS) was therefore chosen for generation of DNA methylation profiles from serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs. One biological replicate was sequenced for each cell line/condition (six samples in total). Library production and Illumina sequencing was performed by the Epigenomics Core Facility of Weill Cornell Medicine (New York, USA), while the data was analysed bioinformatically by Dr Donncha Dunican.

4.2.2.1 The DNA methylomes of lsh$^{-/-}$ mESCs are distinct from those of E14 mESCs

The transcriptional profiles of serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs were shown to cluster based on conditions in which the mESCs were cultured in, rather than on the genotype in relation to *Lsh*. To determine whether this was the case for the DNA methylation profiles of these samples, a similar HCA was performed using the ERRBS datasets. This revealed that E14 and *lsh*$^{-/-}$ mESC DNA methylomes initially cluster based on *Lsh* genotype, with serum and reversion *lsh*$^{-/-}$ mESCs pairing up with one another and serum and reversion E14 mESCs pairing up separately (Figure 4.4 A). This demonstrates that the DNA methylomes of serum and reversion mESCs are similar within each cell line, indicating that during reversion from the hypomethylated ground state, 5-mC patterns are generally re-established to the genomic loci that exhibited methylation prior to 2i adaptation. The fact that *lsh*$^{-/-}$ and E14 mESCs cluster separately, in contrast to the transcriptome data, suggests that the DNA methylation profiles of *lsh*$^{-/-}$ mESCs cultured in serum are distinct from those of the equivalent E14 mESCs. However, although there is some degree of distinction between E14 and *lsh*$^{-/-}$ mESCs cultured in conditions containing serum, all these samples are present on the same branch of the dendogram, indicating that their 5-mC profiles remain closely related. As with the gene expression profiles shown in Figure 4.1 A, the DNA methylomes of 2i E14 and *lsh*$^{-/-}$ mESCs cluster away from those of the mESCs cultured in serum-containing conditions. However, 2i E14 and *lsh*$^{-/-}$ mESCs do not cluster on

the same branch as each other, suggesting there is also a distinction between the DNA methylomes of ground state mESCs in the presence and absence of *Lsh*.

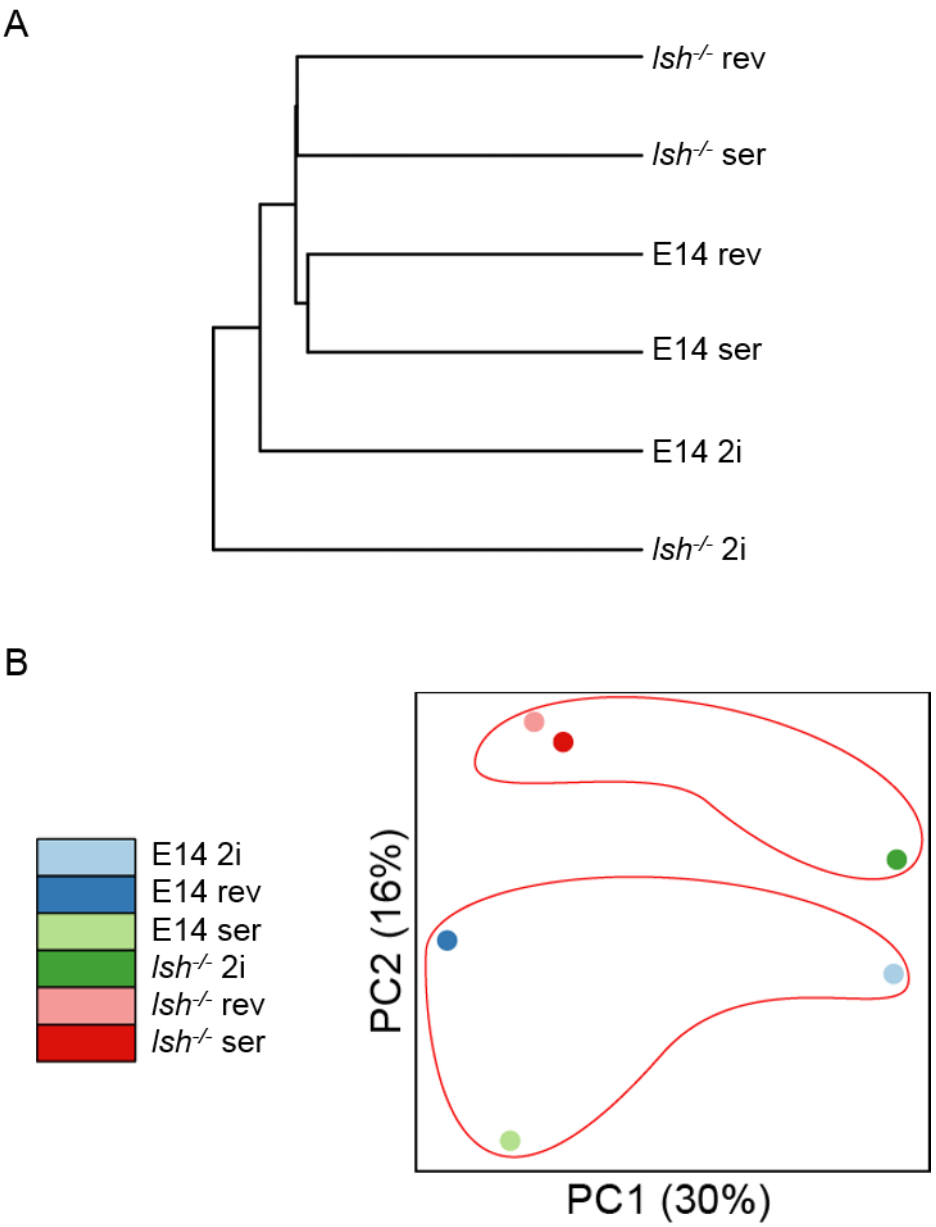


Figure 4.4. The DNA methylomes of *lsh*$^{-/-}$ mESCs are distinct from those of E14 mESCs. Analysis of the degree of similarity between serum, 2i and reversion E14 and *lsh*$^{-/-}$ ERRBS datasets using HCA (A) and PCA (B).

To further assess the differences between the DNA methylomes of serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs, PCA was performed. This demonstrated that the main source of variation (30%) between the ERRBS datasets was due to the differences between mESCs cultured in conditions containing serum or 2i, indicating that the 5-

mC profiles of E14 and *lsh*$^{-/-}$ mESCs do initially cluster based on culture conditions (Figure 4.4 B). However, 16% of the variation between the DNA methylation profiles was accounted for by dissimilarities between E14 and *lsh*$^{-/-}$ mESCs cultured in serum, supporting the observation from the HCA that the DNA methylomes of E14 and *lsh*$^{-/-}$ mESCs cultured in serum are distinct. It also reinforces the suggestion that DNA methylation is re-established in patterns resembling those observed prior to conversion to 2i during reversion from the hypomethylated 2i pluripotent state.

In summary, these analyses illustrate that the DNA methylation profiles of mESCs cultured in serum or 2i are distinct, in concordance with the transcriptome analysis. However, unlike the analysis of gene expression profiles, evaluation of the DNA methylomes E14 and *lsh*$^{-/-}$ mESCs highlights differences in 5-mC profiles between these mESC lines, particularly when they are cultured in conditions containing serum.

4.2.2.2 Global DNA methylation profiles of E14 and lsh$^{-/-}$ mESCs cultured in serum or 2i are interconvertible

To further examine the genome-wide changes in DNA methylation between serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs, the percentage gain or loss of methylation at individual CpGs between samples was calculated and collated in a violin plot to indicate changes in global DNA methylation levels (Figure 4.5 A). This clearly illustrates the substantial loss of CpG methylation that occurs in 2i mESCs compared to serum for both E14 and *lsh*$^{-/-}$ mESCs. However, when reversion mESCs are compared to serum, there is very little change in CpG methylation levels in both E14 and *lsh*$^{-/-}$ mESCs, reinforcing the suggestions from the previous analyses that the DNA methylation profiles of serum and reversion mESCs are very similar in both E14 and *lsh*$^{-/-}$ backgrounds. The comparison of *lsh*$^{-/-}$ mESCs with E14 in a serum or 2i context also demonstrates that there appears to be relatively little gain or loss in overall CpG methylation in the absence of *Lsh*.

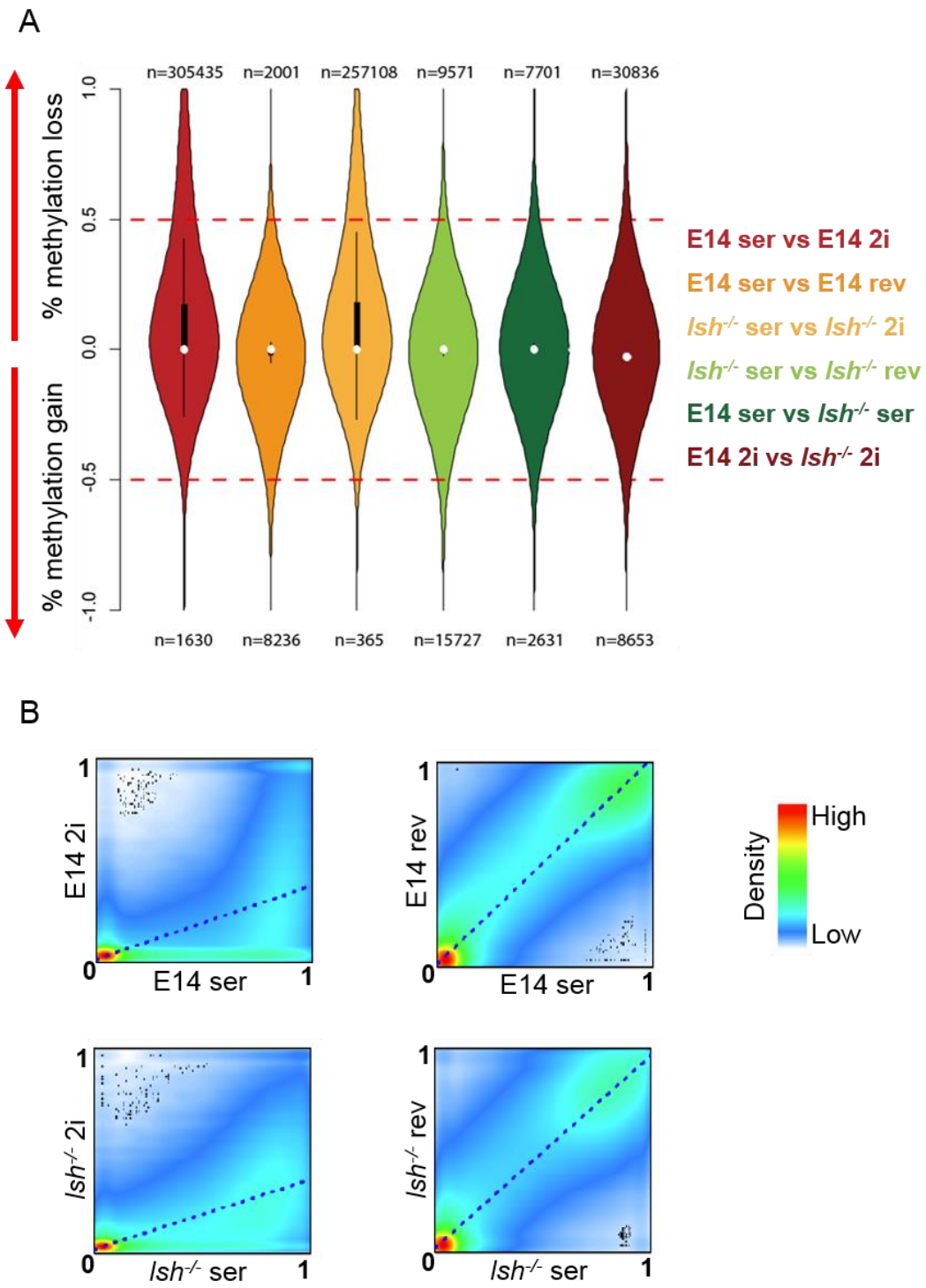


Figure 4.5. Global DNA methylation profiles of E14 and *lsh*$^{-/-}$ mESCs cultured in serum or 2i are interconvertible. Analysis of serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESC ERRBS datasets to assess changes in methylation of individual CpGs. A. Violin plot showing individual CpGs that gain or lose methylation between samples indicated on the right hand side. The scale 0-1 represents 0-100% methylation gain/loss. Number of CpGs (n) that lose or gain more than or equal to 50% methylation (indicated by the red dotted line) are shown above and below each plot, respectively. Black bars on the plots show the interquartile range. B. Scatter plots comparing percentage methylation of

individual CpGs within the samples indicated on the axes. The scale 0-1 represents 0-100% CpG methylation. Heatmap on each plot is generated using two-dimensional kernel density estimation and represents the density of data points at that percentage of CpG methylation within each sample. Black dots identify outlier data points from areas of lowest regional density.

The DNA methylation profiles of serum, 2i and reversion mESCs were compared further using a scatter plot and a two-dimensional kernel density estimation. The line of best fit was determined using the linear least squares regression method. This provided a further clear demonstration of the global loss in DNA methylation that occurs following adaptation to 2i in both E14 and *lsh*$^{-/-}$ mESCs, as the higher density of data points along the x-axes indicates increased DNA methylation in serum E14 (upper panels) and *lsh*$^{-/-}$ mESCs (lower panels, Figure 4.5 B). Furthermore, the line of best fit is skewed towards the x-axis in both scatter plots, which emphasises the increased density of data points along this axis, indicating a higher occurrence of methylated CpGs in the serum mESCs compared to 2i for both E14 and *lsh*$^{-/-}$ mESCs. When serum mESCs are plotted against reversion mESCs, the almost perfectly diagonal line of best fit demonstrates the similarity of the DNA methylation profiles between these samples for both E14 and *lsh*$^{-/-}$ mESCs. Along with the violin plots, this clearly shows that the 2i-associated global hypomethylation is recovered following reversion into serum, resulting in reversion mESCs with similar 5-mC profiles to those observed prior to 2i adaptation. This demonstrates the interconvertibility between the DNA methylomes of mESCs transitioned between serum and 2i culture. Furthermore, these analyses show no notable difference in the degree of interconvertibility between E14 and *lsh*$^{-/-}$ mESCs, suggesting that lack of *Lsh* has little impact on the ability of mESCs to undergo DNA methylome remodelling during transitions between *in vitro* pluripotent states.

4.2.2.3 Ground state lsh$^{-/-}$ mESCs exhibit moderate changes in DNA methylation compared to 2i E14 mESCs

To investigate in more detail whether there were any subtle differences in the DNA methylomes of E14 and *lsh*$^{-/-}$ mESCs, the 5-mC profiles of these samples were compared using the scatter plots described previously. In the violin plots displayed in Figure 4.5 A, a modest number of CpGs were shown gain and lose methylation in

serum *lsh*$^{-/-}$ mESCs compared to E14 (2631 and 7701, respectively). These changes were apparent when serum E14 mESCs were plotted against serum *lsh*$^{-/-}$ mESCs, manifesting in a small skew of the line of best fit towards serum E14 mESCs (Figure 4.6 A). This indicates the presence of slightly more methylated CpGs in the serum E14 mESC dataset.

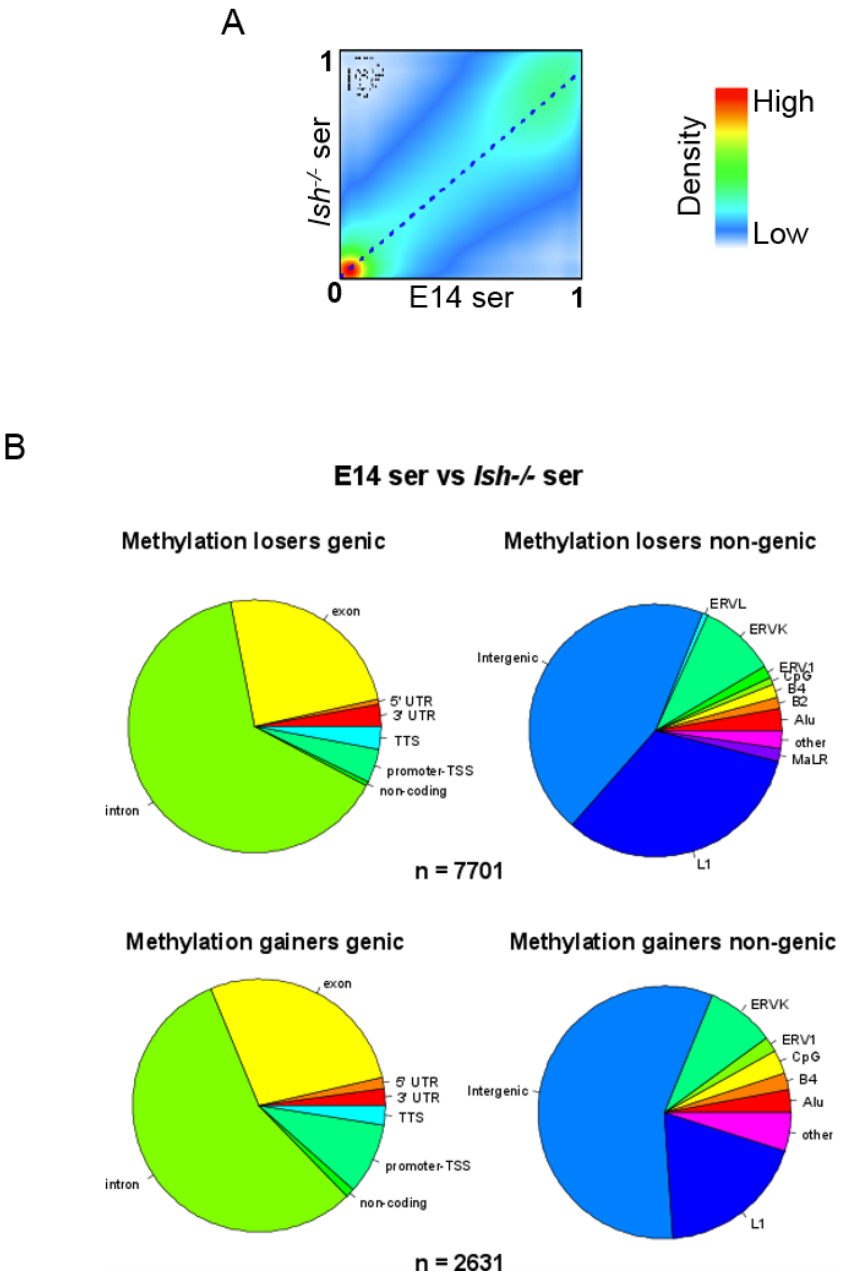


Figure 4.6. A modest number of CpGs exhibit changes in 5-mC status in *serum lsh*$^{-/-}$ mESCs compared to serum E14 mESCs. A. Scatter plot with two-dimensional kernel density comparing CpG methylation of serum E14 mESCs to serum *lsh*$^{-/-}$ mESCs. B. Analysis of genic and non-genic regions in which CpGs that lose/gain methylation between serum *lsh*$^{-/-}$ and E14 mESCs (identified in Figure 4.5 A) reside.

To assess the genomic loci affected by loss or gain of DNA methylation in serum *lsh*$^{-/-}$ mESCs, the regions in which the CpGs identified in the analysis shown in Figure 4.5 A reside were determined and classified into genic and non-genic regions (using mm9 as a reference genome). CpGs that lose methylation in serum *lsh*$^{-/-}$ mESCs compared to E14 are enriched in intronic regions, with exons also being affected to a lesser extent (Figure 4.6 B, left panels). Other genomic features such as promoters, transcription termination sites (TTS) and 3' and 5' untranslated regions (UTRs) are also affected to a much lesser extent. A similar pattern of affected genic regions is observed for CpGs that gain methylation in serum *lsh*$^{-/-}$ mESCs. The non-genic regions affected by CpG methylation gain or loss are also similar to one another, with intergenic sequences being enriched the most, followed by LINE-1 elements, then a range of other retroviral elements, such as ERVs, affected to a lesser extent (Figure 4.6 B, right panels). This suggests that *Lsh* modulates DNA methylation mostly at a small subset of introns, exons, intergenic and LINE-1 sequences in serum *lsh*$^{-/-}$ mESCs.

To explore whether a lack of *Lsh* impacts the levels or genomic distribution of DNA methylation in the pluripotent ground state, the same analysis was performed comparing 2i *lsh*$^{-/-}$ mESCs to 2i E14 mESCs. Interestingly, the scatter plot showed a striking skew of the line of best fit towards the 2i E14 mESCs, indicating that these mESCs contained considerably more methylated CpGs than 2i *lsh*$^{-/-}$ mESCs (Figure 4.7 A). Indeed, the violin plot analysis shown in Figure 4.5 A identified a moderate number of CpGs that gained and, in particular, lost cytosine methylation in 2i *lsh*$^{-/-}$ mESCs compared to E14 (8653 and 30836, respectively). Establishment of the genomic regions that these CpGs reside in revealed a similar pattern to the serum *lsh*$^{-/-}$ mESCs, with introns comprising more than half of the affected genic regions, followed by exons constituting approximately a quarter (Figure 4.7 B, left panels). Non-genic regions mostly affected by CpG methylation gain in 2i *lsh*$^{-/-}$ mESCs also mimicked serum *lsh*$^{-/-}$ mESCs, with intergenic loci being the primary target for aberrant CpG methylation, followed by LINE-1 elements. Examination of the loci encompassing CpGs which lost cytosine methylation in 2i *lsh*$^{-/-}$ mESCs revealed enrichment of intergenic sequences and regions classified as 'other'. However, it also showed that almost half of the non-genic regions in which the CpGs that lose methylation reside were LINE-1 subfamilies, suggesting that *Lsh* plays a role in

establishing or maintaining DNA methylation at these repetitive elements in the ground state of pluripotency.

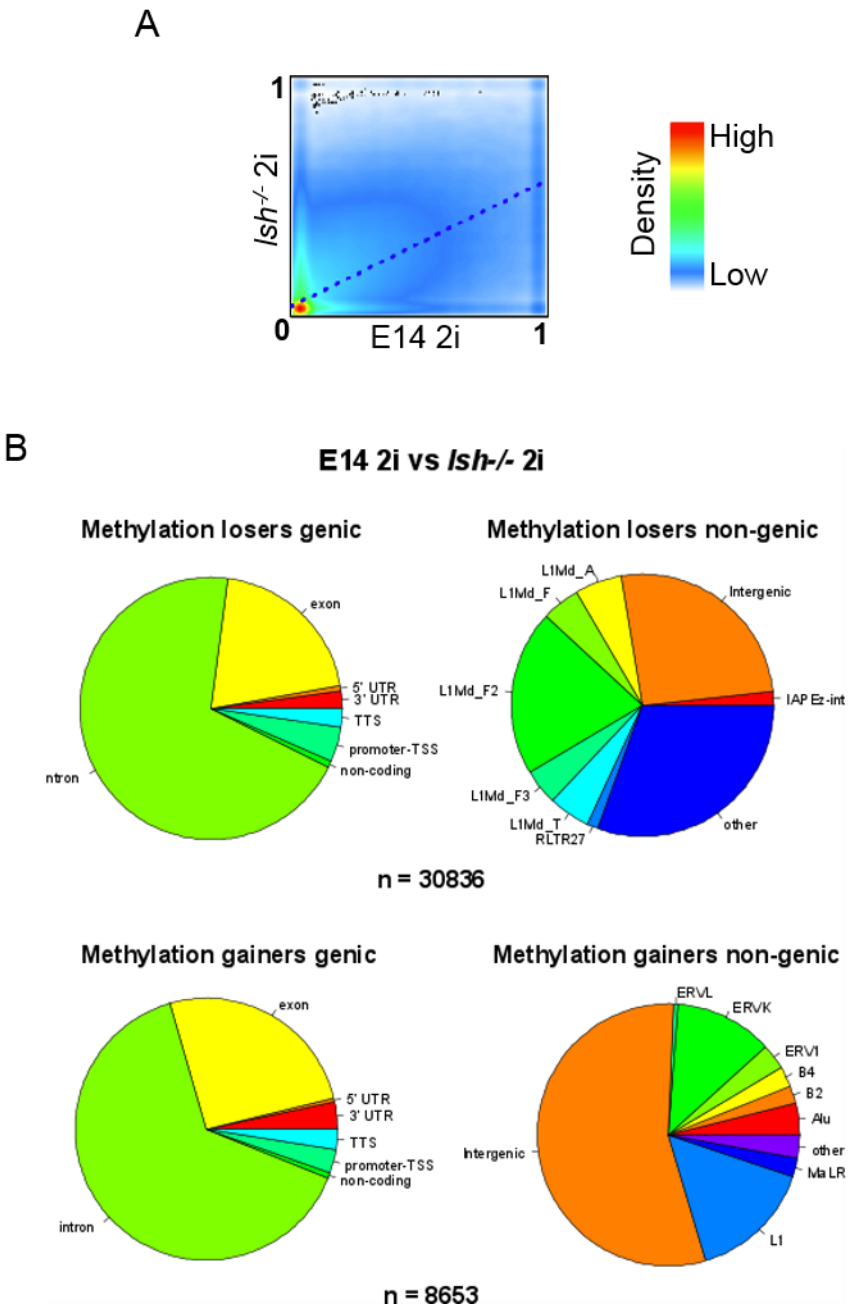


Figure 4.7. A moderate number of CpGs lose and gain methylation in 2i *lsh*$^{-/-}$ mESCs compared to 2i E14 mESCs. A. Scatter plot with two-dimensional kernel density comparing CpG methylation of 2i E14 mESCs to 2i *lsh*$^{-/-}$ mESCs. B. Analysis of genic and non-genic regions in which CpGs that lose/gain methylation between 2i *lsh*$^{-/-}$ and E14 mESCs (identified in Figure 4.5 A) reside.

4.3 Discussion

The bioinformatic analysis of high-resolution RNA-seq and ERRBS datasets described in this chapter was undertaken to examine the genome-wide changes in gene expression and DNA methylation that occur during transition between serum and 2i culture in E14 and *lsh*$^{-/-}$ mESCs. These analyses demonstrated that the transcriptomes and, even more so, DNA methylomes of mESCs are largely interconvertible between the two pluripotent states. Furthermore, it showed that while the absence of *Lsh* may have a subtle effect on the transition of gene expression patterns between these two culture states, it does not appear to have any notable effect on the interconvertibility of serum and 2i mESC DNA methylation profiles. However, it was noticed that the DNA methylomes of *lsh*$^{-/-}$ mESCs cultured in serum are distinct from serum E14 mESCs, and that there is a degree of DNA hypomethylation in 2i *lsh*$^{-/-}$ mESCs, a proportion of which occurs at LINE-1 element subfamilies. This suggests that while *Lsh* may not play a role in the global remodelling of DNA methylation patterns during transitions between serum and 2i culture, its absence does result in modest perturbations of 5-mC profiles in both serum and 2i mESCs.

4.3.1 Transcriptional and DNA methylation profiles are generally interconvertible between serum and 2i culture

In Chapter 3, I demonstrated that the serum and 2i culture states appear to be interconvertible in terms of colony morphology, expression of a few key genes and global DNA methylation levels. However, the transcriptional changes that occur between the two pluripotent states were only assessed for a few genes, while only total levels of 5-mC were measured in serum and 2i mESCs. I aimed to evaluate whether these smaller-scale changes were applicable genome-wide through examination of transcription and DNA methylation profiles of serum, 2i and reversion mESCs. Scatter plot analysis of gene expression profiles generated by RNA-seq demonstrate that a vast number of genes are upregulated and downregulated more than two-fold upon transition to 2i, as had been reported previously (Marks *et al.*, 2012). Following reversion into serum, these expression changes are largely reversed, resulting in a transcriptional profile that is mostly comparable to that of serum mESCs prior to 2i adaptation. Some differences in gene expression profiles were apparent, suggesting

that transcription patterns do not completely return to those observed initially in serum. However, the gene expression trends generally follow the pattern demonstrated by the candidate genes assessed in Chapter 3 during transition between serum and 2i culture. PCA and scatter plot analysis also clearly showed that the DNA methylomes of reversion mESCs return to profiles highly similar to those observed initially in serum mESCs. Indeed, HCA revealed that the DNA methylomes of serum and reversion mESCs cluster together, indicating the high degree of similarity between the 5-mC profiles. This demonstrates that not only do bulk levels of 5-mC recover following 2i-associated DNA hypomethylation, the patterns of 5-mC also return to those observed prior to 2i adaptation, confirming the findings from previous reports (Habibi *et al.*, 2013; Leitch *et al.*, 2013; Choi *et al.*, 2017). Together, these results provide further evidence for the faithful interconversion of transcription and DNA methylation profiles between the pluripotent serum and 2i culture states.

4.3.2 Culture conditions have more impact on mESC transcriptomes and DNA methylomes than absence of Lsh

The expression profiles and DNA methylomes of E14 and *lsh*$^{-/-}$ mESCs both undergo extensive changes upon adaptation to and reversion from 2i culture, as demonstrated by the transcription and DNA methylation profiling analyses shown in this chapter. This observation is further demonstrated by comparison of serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs transcription and 5-mC profiles using HCA and PCA. These analyses showed that both E14 and *lsh*$^{-/-}$ mESCs clustered closely together depending on the conditions in which they were cultured, with the principal component of variation being the difference between serum and 2i profiles for both the gene expression and DNA methylation analysis. This shows that the conditions in which mESCs are cultured is the primary modulator of genome-wide transcription and DNA methylation patterns, rather than the genotype of the mESCs in relation to *Lsh*. The DNA methylome PCA did however show that the secondary contributor to the variation observed between ERRBS datasets was accounted for by differences between E14 and *lsh*$^{-/-}$ mESCs cultured in serum. This implies that the absence of *Lsh* does in fact have an impact on DNA methylation patterns in mESCs, resulting in *lsh*$^{-/-}$ mESC 5-mC profiles that are distinct from those of E14 mESCs, with this distinction

particularly evident in mESCs cultured in the presence of serum. This finding was alluded to in the LC-MS global DNA methylation analysis of serum, 2i and reversion E14 and *lsh*$^{-/-}$ mESCs in Chapter 3, in which the total 5-mC levels of serum *lsh*$^{-/-}$ mESCs appeared to be lower than in E14, although this difference was not found to be significant.

4.3.3 Lack of Lsh results in gene expression and DNA methylation profile perturbations

Although the absence of *Lsh* did not have any profound effects on the extensive gene expression changes that occur during transition to or reversion from 2i, a small subset of genes were found to be mis-expressed in serum *lsh*$^{-/-}$ mESCs compared to serum E14 mESCs. It was also noted that a large proportion of these genes exhibited a similar change in expression in *lsh*$^{-/-}$ mESCs to the change in expression that is observed during 2i adaptation in E14 mESCs, suggesting that *Lsh* influences the expression of a selection of genes involved in transition to the ground state of pluripotency. GO analysis of these mis-expressed genes revealed an enrichment in terms associated with signalling mechanisms and sensory perception, providing insight into the general function of genes that are mis-expressed in the absence of *Lsh* in a pluripotent context.

Since PCA of DNA methylomes highlighted differences in 5-mC profiles between E14 and *lsh*$^{-/-}$ mESCs cultured in serum, the extent of these perturbations and the genomic regions affected by DNA methylation gain or loss in the absence of *Lsh* were assessed. This revealed a modest number of CpGs affected by differences in cytosine methylation in serum *lsh*$^{-/-}$ mESCs compared to E14, and that the principal genomic regions affected were widespread, not concentrated in any specific genome feature. Comparison of the DNA methylation profiles of 2i *lsh*$^{-/-}$ mESCs with 2i E14 mESCs revealed a greater number of CpGs where methylation was gained or, principally, lost. Interestingly, a proportion of CpGs that lost methylation were located in LINE-1 element subfamilies, suggesting that *Lsh* contributes to the establishment or maintenance of DNA methylation at these repetitive elements in ground state mESCs. This is consistent with previous reports demonstrating LINE-1 hypomethylation in *Lsh*-depleted mouse embryonic tissues and somatic cells (Dennis *et al.*, 2001; Dunican *et al.*, 2013; Ren *et al.*, 2015).

In Chapter 3 it was noted that serum *lsh*$^{-/-}$ mESCs displayed some similar characteristics to 2i E14 mESCs, such as expression of a few DNMTs and lineage-associated factors. Therefore, it was proposed that serum *lsh*$^{-/-}$ mESCs could represent an intermediate state between serum and 2i. Comparison of genome-wide transcriptional profiles of serum *lsh*$^{-/-}$ and 2i E14 mESCs did indeed identify a small subset of genes that exhibited similar expression between these two mESC lines (Figure 4.3 A). However, HCA and PCA has revealed that, generally, serum *lsh*$^{-/-}$ mESCs do not appear to represent an intermediate state between serum and 2i in terms of transcription and DNA methylation. However, there are perturbations in patterns of expression and, in particular, DNA methylation in serum and 2i *lsh*$^{-/-}$ mESCs compared to WT, suggesting that *Lsh* does indeed play a role in modulating DNA methylation patterns in a pluripotent context.

4.3.4 Technical considerations for DNA methylome analysis by ERRBS

The profiling of DNA methylation in E14 and *lsh*$^{-/-}$ mESCs using ERRBS revealed a moderate amount of changes in CpG methylation between these mESC lines cultured in serum and 2i. Despite the coverage of ERRBS purportedly being double that of RRBS, allowing the examination of exons and introns, the majority of the sequencing reads remain enriched in genomic regions containing high densities of CpGs, such as CGIs and gene promoters (Garrett-Bakelman *et al.*, 2015). Therefore, the full extent of DNA methylation differences in *lsh*$^{-/-}$ mESCs compared to E14 may not be apparent using the ERRBS technique. In particular, the coverage of relatively CpG-poor (in comparison to CGIs) genomic regions such as repetitive elements is low using this method. Previous studies have shown repetitive regions of the genome to be the primary target of *Lsh* (Dennis *et al.*, 2001; Huang *et al.*, 2004; Dunican *et al.*, 2013; Yu, McIntosh, *et al.*, 2014). Therefore, further analysis of the DNA methylomes of *lsh*$^{-/-}$ mESCs using an unbiased method such as WGBS would be ideal for further elucidation of the impact of *Lsh* deletion on genome-wide DNA methylation patterns.

Another caveat of the ERRBS method, and other methods using bisulfite treatment to identify methylated cytosines, is that bisulfite treatment cannot distinguish between 5-mC and 5-hmC. Global levels of 5-hmC have been shown to increase during the initial stages of transition from serum to 2i culture (Ficz *et al.*, 2013; Habibi *et al.*, 2013).

However, these elevated 5-hmC levels decrease to a lower, more stable level after 72 hours of 2i adaptation (Ficz *et al.*, 2013). Therefore, although the fact that these two modified bases cannot be distinguished should be kept in mind, the much lower genomic levels of 5-hmC compared to 5-mC mean that the vast majority of each DNA methylation profile generated using ERRBS will accurately represent the patterns of 5-mC in the genome.

Chapter 5. Lsh is required for establishment of DNA methylation in epiblast stem cells and embryoid bodies

5.1 Introduction

In Chapter 3 I demonstrated that *Lsh* does not appear to contribute to DNA methylation re-establishment during transition from the hypomethylated ground state of pluripotency (2i) to a more methylated, primed state of pluripotency (serum). However, *Lsh* has convincingly been shown to be required for DNA methylation globally and at specific genomic loci in mouse embryonic tissue and somatic cells (Dennis *et al.*, 2001; Sun, David W. Lee, *et al.*, 2004; Myant *et al.*, 2011; Dunican *et al.*, 2013; Yu, McIntosh, *et al.*, 2014). This suggests that *Lsh* is required for establishment or maintenance of DNA methylation patterns during embryonic development. Most reports point to a role for *Lsh* in *de novo* DNA methylation through its association with the *de novo* methyltransferases *Dnmt3a* and *Dnmt3b* (Zhu *et al.*, 2006; Myant and Stancheva, 2008; Ren *et al.*, 2015; Termanis *et al.*, 2016). Therefore, *Lsh* may contribute to the wave of *de novo* DNA methylation that occurs in the early stages of development during implantation of the blastocyst (Borgel *et al.*, 2010; Smith *et al.*, 2012). However, previous investigations of *Lsh* function during the early stages of development are limited, with most studies having concentrated on the impact of *Lsh* depletion at later developmental stages, in somatic cells and tissues.

In this chapter, I aimed to address the requirement for *Lsh* in establishing DNA methylation during transition from ground state to post-implantation development, beyond the 'primed' pluripotent state that serum mESCs supposedly reflect. I used two culture systems to model post-implantation development – embryoid bodies (EBs) and epiblast stem cells (EpiSCs) – to increasingly narrow down the developmental window in which *Lsh* acts to influence DNA methylation. I examined the impact of absence of *Lsh* on global DNA methylation in these culture models, as well as on centromeric and pericentromeric repetitive elements, to deduce the contribution of *Lsh* to *de novo* DNA

methylation globally and locally during the early stages of post-implantation development that these culture models represent.

5.2 Results

5.2.1 Differentiation of 2i mESCs to EBs

5.2.1.1 Experimental strategy

To dissect the requirement for *Lsh* in facilitating DNA methylation during the differentiation process *in vivo*, I aimed to differentiate E14 and *lsh*$^{-/-}$ mESCs to EBs *in vitro*. A report from Ren *et al*. (2015) demonstrates that *Lsh* is required for *de novo* DNA methylation of IAP, LINE-1 and minor satellite repeats during differentiation to EBs in the presence of RA, which directs differentiating mESCs down a neuronal lineage. I sought to confirm and extend this analysis to examine the contribution of *Lsh* to global DNA methylation levels in the presence and absence of RA, as well as evaluating the effect of *Lsh* deletion on DNA methylation levels locally at major and minor satellites. To confirm successful differentiation of E14 and *lsh*$^{-/-}$ mESCs to EBs, changes in cell morphology and expression of key lineage-specific genes were assessed (outlined in Figure 5.1). The impact of absence of *Lsh* on DNA methylation globally and at major and minor satellites was monitored in EBs differentiated in the presence and absence of RA to delineate whether *Lsh* contributes to DNA methylation only during differentiation down the neuronal lineage, or whether it functions more generally during generation of all three germ layers.

As discussed previously, serum mESCs exhibit a high level of global DNA methylation similar to that of differentiated cells, while 2i mESCs are globally hypomethylated (Senner *et al.*, 2012; Habibi *et al.*, 2013; Leitch *et al.*, 2013). Therefore, I reasoned that beginning the differentiation protocol with 2i mESCs would allow easier examination of the ability of E14 and *lsh*$^{-/-}$ mESCs to re-establish global DNA methylation patterns during differentiation to EBs due to the substantial difference in bulk 5-mC levels between these culture models. Furthermore, since the methylomes of 2i mESCs are proposed to more faithfully reflect those of the ICM, it could be argued that using ground state mESCs as a starting point represents a more physiologically relevant model of the DNA methylation establishment that occurs

during differentiation from a naïve pluripotent state to a developmentally restricted state. Therefore, I utilised E14 and *lsh*$^{-/-}$ mESCs that had been adapted to 2i for 15 days (six passages) as a starting point for the differentiation to EBs in the presence and absence of RA (experimental strategy depicted in Figure 5.1).

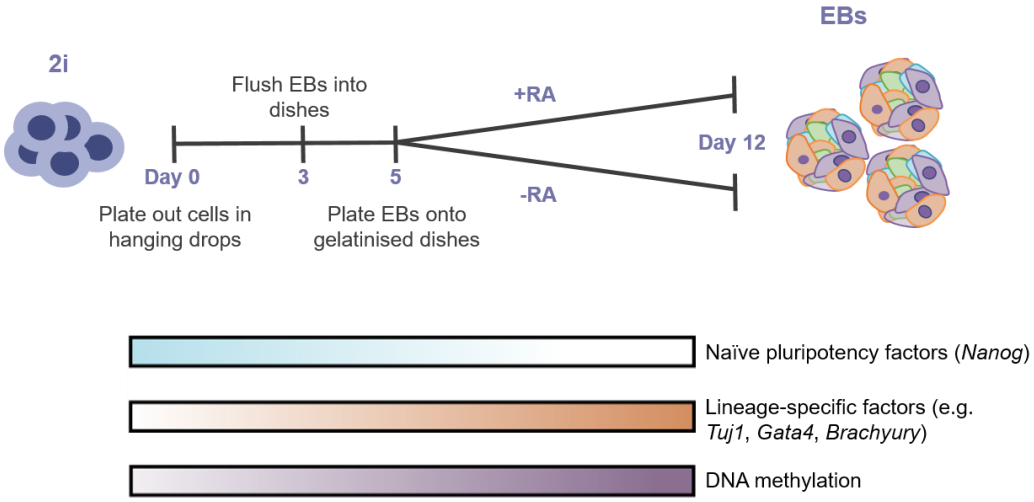


Figure 5.1. Diagram depicting the experimental workflow of the differentiation of 2i mESCs to EBs in the presence/absence of RA. Experimental procedures for generation of EBs are summarised, from plating out 2i mESCs in hanging drops at day 0 to harvesting EBs at day 12. Key changes in pluripotency and lineage marker expression and 5-mC levels are highlighted.

5.2.1.2 E14 and lsh$^{-/-}$ mESCs adapt morphologically during differentiation to EBs

The morphology of differentiated cells is vastly different to that of ground state mESCs. As reported by Marks and Stunnenberg (2014) and demonstrated in Chapter 3, mESCs adapted to 2i form compact, three-dimensional colonies that are rounded with smoothened colony boundaries (Figure 3.4). In contrast, differentiated cells adopt a vast array of morphologies dependent on cell type, and tend to grow in monolayers of flattened, irregularly-shaped cells *in vitro*. Following aggregation of 2i mESCs to form EB structures, the EBs were allowed to adhere and the outgrowth was monitored to assess the generation of differentiated cells displaying a range of morphologies from E14 and *lsh*$^{-/-}$ 2i mESCs.

E14 and *lsh*$^{-/-}$ 2i mESCs exhibited typical 2i morphology prior to differentiation, as described previously (Figure 5.2, left panels). Following differentiation to EBs, a

monolayer of cells exhibiting a variety of morphologies was evident in E14 and *lsh*$^{-/-}$ cultures in the presence and absence of RA, indicating that both E14 and *lsh*$^{-/-}$ mESCs had successfully differentiated into a range of cell types (Figure 5.2, middle and right panels). As expected, cultures generated in the presence of RA exhibited less morphological heterogeneity and increased presence of cells displaying neuronal-like projections (indicated by the arrows in Figure 5.2, right panels). This suggested that both E14 and *lsh*$^{-/-}$ EBs supplemented with RA resulted in cell populations largely committed to the neuronal lineage. Although analysis of lineage-specific gene expression is required to confirm generation of differentiated cultures representing all three germ layers, these morphological changes from 2i mESCs to EBs provided an initial indication that both E14 and *lsh*$^{-/-}$ mESCs are able to generate differentiated cultures, and that the presence of RA promotes generation of neuron-like cells.

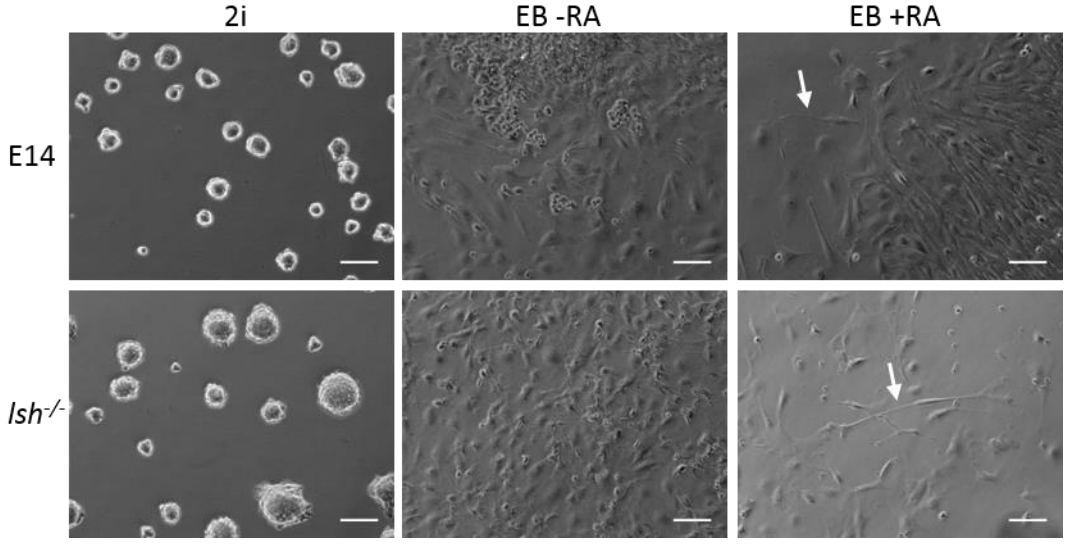


Figure 5.2. E14 and *lsh*$^{-/-}$ mESCs both differentiate to EBs in the presence and absence of RA. Brightfield images showing morphological changes that occur during differentiation of E14 and *lsh*$^{-/-}$ mESCs cultured in 2i (left panels) to EBs in the absence and presence of RA in the middle and right panels, respectively. Arrows highlight the presence of cells with neural-like projections in EBs cultured in the presence of RA. Scale bars represent 100 µm.

5.2.1.3 E14 and lsh$^{-/-}$ *mESCs transcriptionally upregulate lineage-specific markers following differentiation to EBs*

The differentiation of mESCs to EBs is accompanied by the upregulation of lineage-specific genes and the downregulation of naïve pluripotency factors (Murry and Keller,

2008; Radzisheuskaya *et al.*, 2013). The effective differentiation into all three germ layers was judged by the increased expression of markers representing all three embryonic lineages: ectoderm, endoderm and mesoderm. Immunofluorescence and qRT-PCR analysis were used to evaluate the expression of lineage-specific factors in E14 and *lsh*$^{-/-}$ 2i mESCs and EBs generated in the presence and absence of RA.

An increase in the mesodermal marker *α-smooth muscle actin* was clearly visible in E14 and *lsh*$^{-/-}$ EBs in the presence and absence of RA, as demonstrated by immunofluorescence (Figure 5.3 A). In contrast, both E14 and *lsh*$^{-/-}$ mESCs were completely devoid of expression of this marker of the mesodermal lineage. Immunofluorescence for the neuroectodermal-specific *class III β–tubulin* (*Tuj1*) gene demonstrated vast upregulation of this factor in E14 and *lsh*$^{-/-}$ EBs cultured in the presence of RA (Figure 5.3 B). This shows that cultures differentiated in the presence of RA resulted in effective generation of cells committed to the neural lineage. Furthermore, *lsh*$^{-/-}$ EBs exhibited increased *Tuj1* expression in comparison to E14 EBs, suggesting enhanced generation of neuronal cells in the absence of *Lsh*, consistent with a report from Yu *et al.* (2014) that *lsh*$^{-/-}$ iPSCs exhibit an increased propensity to differentiate towards the neural lineage. However, analysis of mRNA levels of the neuroectodermal marker *Nestin* by qRT-PCR did not support this finding, as although *Nestin* expression was greatly upregulated in EBs compared to 2i mESCs, *Nestin* mRNA levels were actually lower in *lsh*$^{-/-}$ EBs cultured in the presence of RA when compared to E14 EBs (Figure 5.4, top left panel). *Nestin* expression was higher in E14 and *lsh*$^{-/-}$ EBs supplemented with RA compared to EBs without RA though, reinforcing the notion that addition of RA promotes differentiation down the neural lineage.

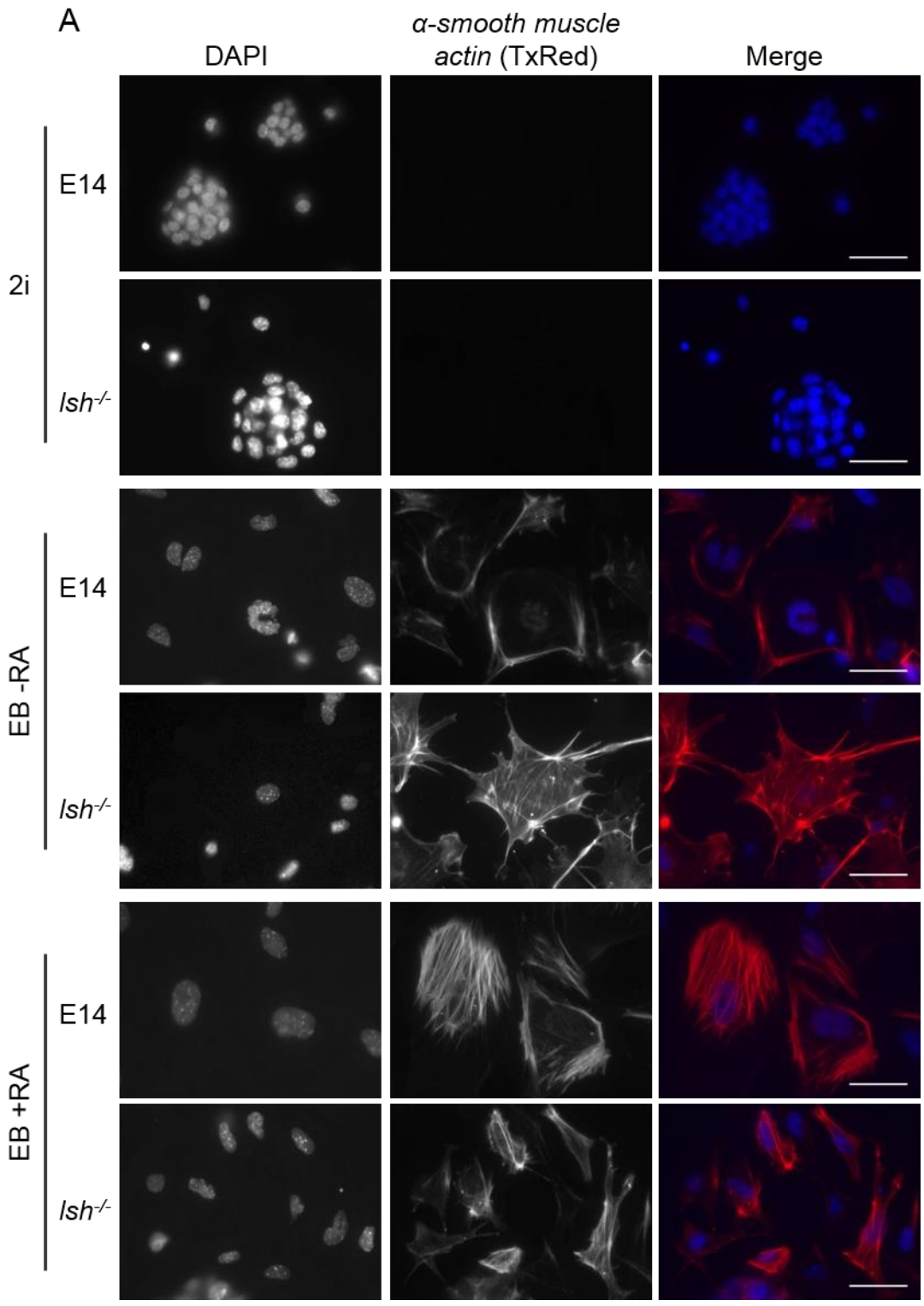
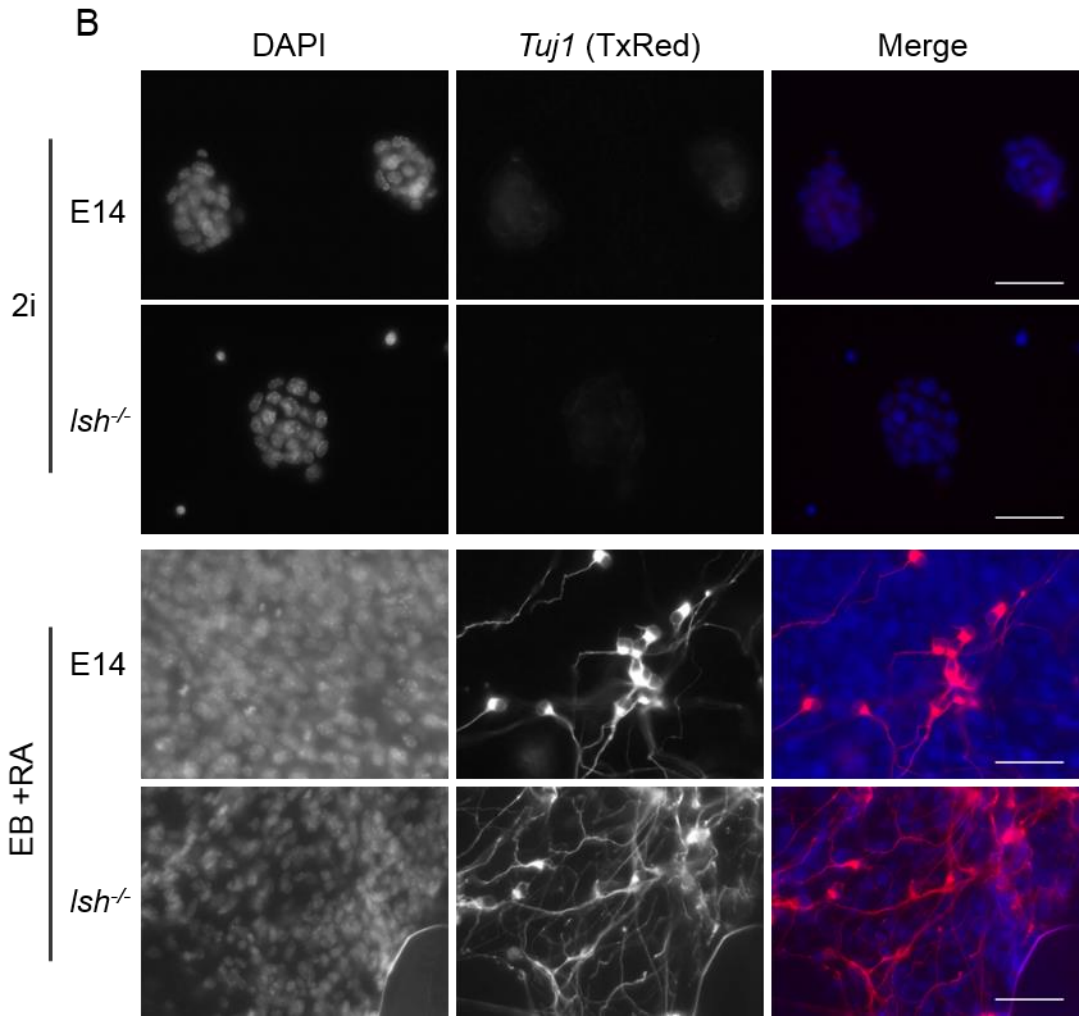


Figure 5.3. E14 and *lsh*$^{-/-}$ mESCs display expression of lineage-associated markers following differentiation to EBs in the presence/absence of RA. Immunofluorescence of E14 and *lsh*$^{-/-}$ 2i mESCs and EBs +/- RA using antibodies against *α-smooth muscle actin* (A) or *Tuj1* (B). Scale bars represent 50 µm.



Further analysis of the expression levels of endodermal and mesodermal markers *Gata4* and *Brachyury* respectively, confirmed that both E14 and *lsh*$^{-/-}$ 2i mESCs produced all three germ layers upon differentiation into EBs. Indeed, *Gata4* mRNA levels substantially increased in EBs generated in the absence of RA (Figure 5.4, top right panel). *Gata4* expression also increased in EBs in the presence of RA, but not to the same extent as without RA, supporting the proposal that EB cultures exhibit increased complexity and heterogeneity in the absence of RA. Interestingly, *Gata4* mRNA levels are notably reduced in *lsh*$^{-/-}$ EBs compared to E14, suggesting that absence of *Lsh* could influence the expression of selected lineage-associated factors. A difference between E14 and *lsh*$^{-/-}$ EBs was also apparent for *Brachyury* expression, which was much increased in *lsh*$^{-/-}$ EBs without RA compared to the equivalent E14

EBs, although a similar pattern was not evident for EBs supplemented with RA (Figure 5.4, bottom left panel). Analysis of further markers from all three germ layers is required to uncover whether a trend exists towards up- or down-regulation of genes belonging to a specific lineage in the absence of *Lsh*. Furthermore, since these results were obtained from qRT-PCR analysis of three technical replicates from one biological replicate for each sample, analysis of additional biological replicates is required to confirm that the observed differences in expression of lineage-associated factors between E14 and *lsh*$^{-/-}$ EBs are reproducible. Aside from this, the vastly increased expression of *Brachyury* in E14 and *lsh*$^{-/-}$ EBs with or without RA confirms the generation of cells representative of the mesodermal lineage. Moreover, examination of *Nanog* mRNA levels demonstrates almost complete transcriptional silencing of this naïve pluripotency marker in both E14 and *lsh*$^{-/-}$ EBs generated in the presence or absence of RA, indicating that these cultures have effectively differentiated to a more developmentally restricted state.

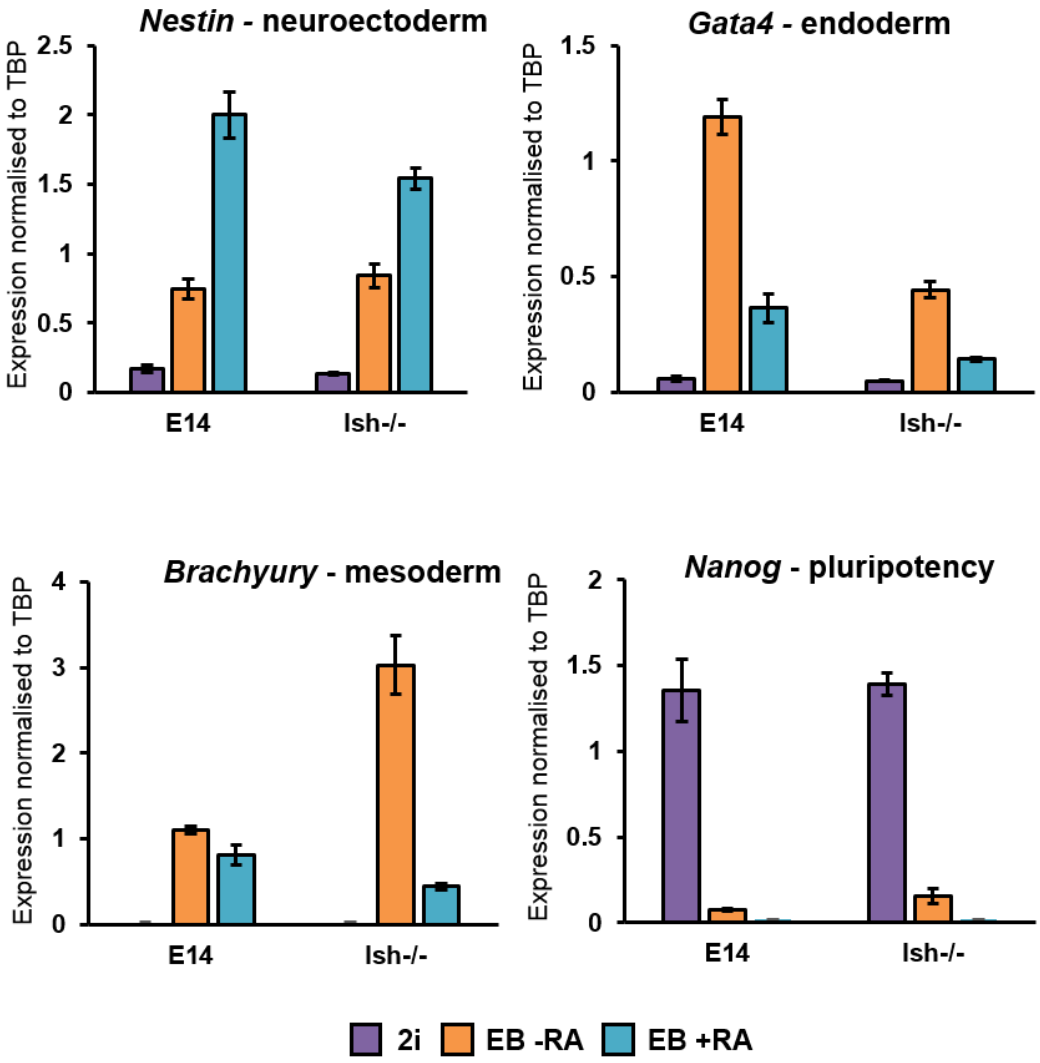


Figure 5.4. E14 and *Ish*$^{-/-}$ cells both show transcriptional changes in key genes following differentiation to EBs in the presence/absence of RA. Quantification of mRNA expression by qRT-PCR of pluripotency or lineage-associated factors (as indicated on graphs), normalised to TBP expression. Error bars represent +/- propagated standard deviation of three technical replicates.

Overall, analysis of these lineage-specific and pluripotency markers by immunofluorescence and qRT-PCR confirmed that both E14 and *Ish*$^{-/-}$ 2i mESCs differentiated to EBs comprised of all three germ layers, but that, as anticipated, EBs generated in the presence of RA largely consist of cells belonging to the neuronal lineage.

5.2.1.4 Global DNA methylation levels are not fully established in *lsh*$^{-/-}$ EBs

To examine the impact of the absence of *Lsh* on global DNA methylation establishment during differentiation of 2i mESCs to EBs, bulk 5-mC levels were measured by LC-MS in E14 and *lsh*$^{-/-}$ 2i mESCs and EBs cultured with and without RA. LC-MS analysis of three technical replicates from one biological replicate for each sample is presented in Figure 5.5. In contrast to the global DNA methylation LC-MS data shown in Figure 3.9, a significant difference of approximately 16% in 5-mC levels between E14 and *lsh*$^{-/-}$ mESCs adapted to 2i is evident (Figure 5.5). A discussion into the potential explanations for the difference between these two datasets is presented in Section 5.3.1. Upon differentiation to EBs, global levels of DNA methylation more than double in both E14 and *lsh*$^{-/-}$ cultures. However, global 5-mC is significantly reduced by approximately 20% in *lsh*$^{-/-}$ EBs generated in the presence and absence of RA when compared to the equivalent E14 EB sample. This significant difference in global DNA methylation levels in *lsh*$^{-/-}$ EBs could be due to defective *de novo* DNA methylation during differentiation to EBs in the absence of *Lsh*. However, this reduction in global 5-mC levels in *lsh*$^{-/-}$ EBs could be accounted for by the difference observed prior to differentiation in 2i mESCs, a possibility that requires further work to fully understand.

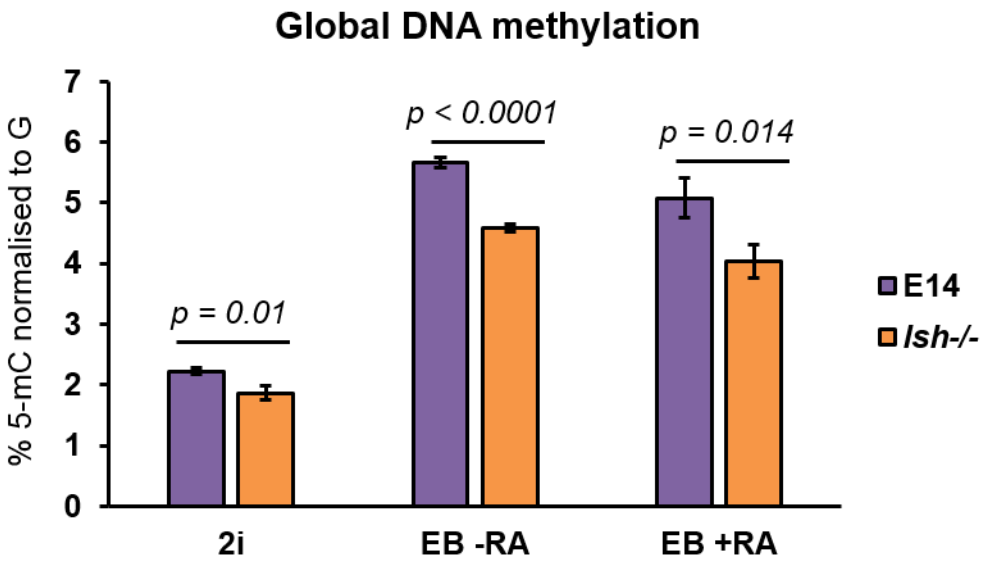


Figure 5.5. Global DNA methylation levels are not fully established in *Ish*$^{-/-}$ EBs differentiated in the presence of RA. Quantification of global 5-mC levels by LC-MS in E14 and *Ish*$^{-/-}$ 2i mESCs and EBs +/- RA. Error bars represent +/- standard deviation of three technical replicates. P-values indicate level of significance between E14 and *Ish*$^{-/-}$ cells as measured by a two-way unpaired t-test.

5.2.1.5 *DNA methylation establishment at major and minor satellites is impaired in the absence of Lsh during differentiation to EBs*

Lsh has convincingly been shown to be required for DNA methylation of a range of repeat classes in mouse embryos and somatic cells (Dennis *et al.*, 2001; Huang *et al.*, 2004; Dunican *et al.*, 2013; Yu, McIntosh, *et al.*, 2014). Recent identification of mutations in the human *Lsh* homologue HELLS as being causative of ICF syndrome, a hallmark of which is hypomethylation at pericentromeric repeats, has highlighted the importance of HELLS in establishing DNA methylation at these repetitive regions (Thijssen *et al.*, 2015). Therefore, I sought to assess the DNA methylation levels at the equivalent repetitive regions in my mESC and EB culture models to further elucidate the role of *Lsh* in DNA methylation of these repeats during differentiation. Initially, I harnessed the convenient feature of methylation-sensitive restriction endonucleases to indicate the cytosine methylation status of CpGs within their associated restriction site through differential restriction digestion, as described in Section 3.2.3.5 (Chapter 3). Again, the occurrence of smaller molecular weight fragments following *MaeII* digestion indicates hypomethylation at its restriction sites, which are enriched in pericentromeric major satellites. Restriction digestion of genomic DNA from E14 and

lsh$^{-/-}$ 2i mESCs and EBs cultured with and without RA revealed similar levels of DNA hypomethylation in E14 and *lsh*$^{-/-}$ 2i mESCs. Following differentiation to EBs, these major satellite regions appear to gain DNA methylation in E14 EBs with and without RA. However, a reduction in DNA methylation at *MaeII* sites was apparent in *lsh*$^{-/-}$ EBs generated in the presence and absence of RA compared to E14, demonstrated by the downwards smear of digested fragments along with the increased presence of lower molecular weight bands (indicated by the arrow in Figure 5.6 A).

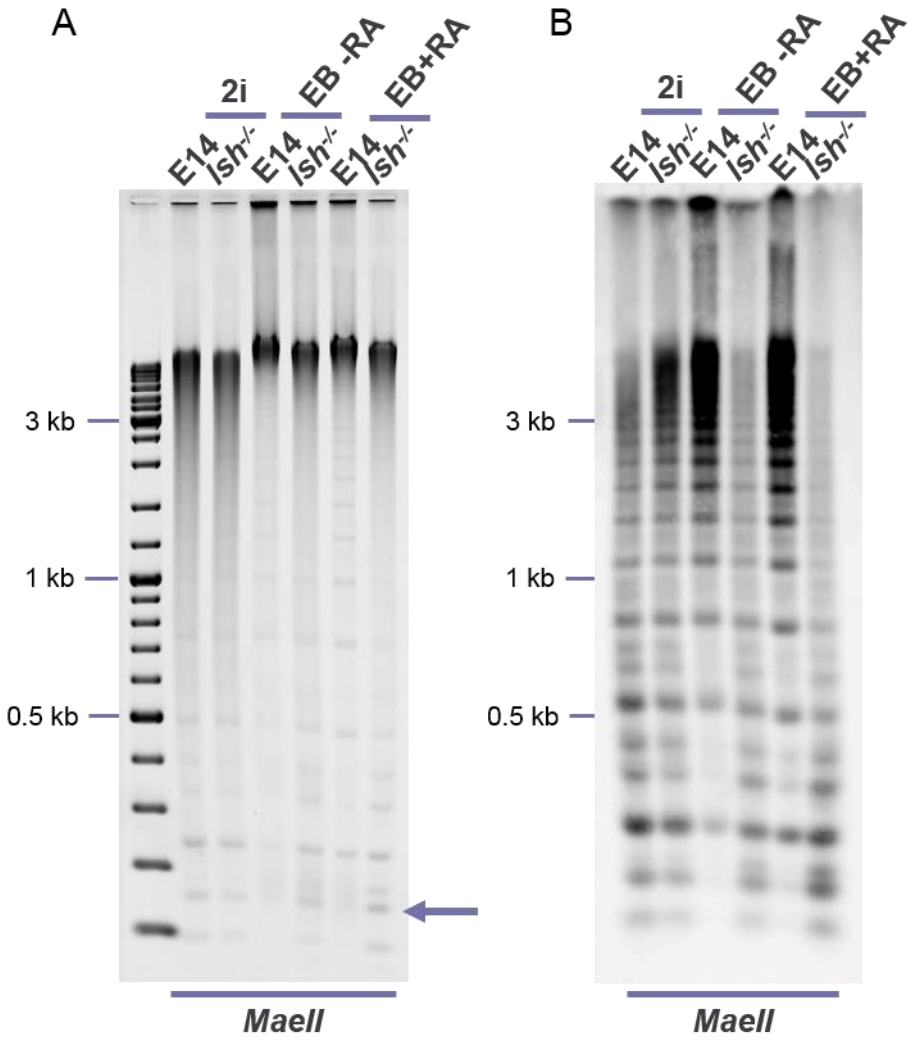


Figure 5.6. DNA methylation is incompletely established at major satellites during differentiation of *lsh*$^{-/-}$ EBs in the presence and absence of RA. A. Agarose gel of E14 and *lsh*$^{-/-}$ 2i, EB –RA and EB +RA *MaeII* restriction digests prior to Southern blot transfer. Arrow highlights smaller molecular weight digested fragments indicating major satellite DNA hypomethylation. B. Southern blot of *MaeII* digest following transfer and hybridised using a probe against major satellites.

To provide extra clarity into the relative levels of major satellite DNA methylation between samples, the analysis using methylation-sensitive digests was combined with a non-radioactive Southern blot using a probe specific for major satellite sequences. The hybridisation of the probe to low molecular weight bands in E14 and *lsh*$^{-/-}$ 2i mESCs indicate that most of the major satellite sequences that the probe is complementary to exhibit some degree of DNA hypomethylation, and are therefore digested by *MaeII* (Figure 5.6 B). The increased signal intensity in the upper portion of the blot in *lsh*$^{-/-}$ 2i mESCs indicates that major satellites exhibit slightly higher 5-mC levels compared to E14. However, following differentiation to EBs in the presence and absence of RA, *lsh*$^{-/-}$ EBs appear to lose some DNA methylation at major satellites, whereas DNA methylation levels in E14 EBs greatly increase, as indicated by the intense staining of digestion-resistant methylated CpGs residing in the upper portion of the blot. This suggests that while major satellites in E14 mESCs gain DNA methylation during differentiation, this *de novo* DNA methylation is impaired in the absence of *Lsh*, which agrees with the analysis of global DNA methylation levels in Figure 5.5.

For analysis of DNA methylation levels at minor satellites, another methylation-sensitive restriction endonuclease, *HpaII*, was utilised as its restriction site (CCGG) is better represented in minor satellite sequences compared to the restriction sites of *MaeII*. *HpaII* digests also provide an estimation of global CpG methylation, as the restriction site is widespread throughout the genome, particularly in regions that are normally hypomethylated, such as CGIs. Digestion with the non-methylation sensitive isoschizomer of *HpaII*, *MspI*, was utilised as a control to demonstrate the equal digestibility of each genomic DNA sample. The *HpaII* digests generally concurred with the global DNA methylation analysis, showing relatively equal DNA hypomethylation of E14 and *lsh*$^{-/-}$ 2i mESCs, as indicated by the general downward shift in the molecular weight of the digested fragments (Figure 5.7 A, left panel). DNA methylation levels were restored following differentiation to EBs in E14 samples, while 5-mC levels appeared to be slightly reduced in *lsh*$^{-/-}$ 2i EBs, agreeing further with the LC-MS analysis. To uncover the relative DNA methylation status of minor satellites between E14 and *lsh*$^{-/-}$ mESCs and EBs, the *HpaII* digests were coupled with a Southern blot using a probe specific for minor satellite sequences. This revealed a

striking reduction of DNA methylation levels in *lsh*$^{-/-}$ 2i mESCs compared to E14 at minor satellite sequences (Figure 5.7 B, left panel). A similar degree of DNA hypomethylation was also apparent in *lsh*$^{-/-}$ EBs cultured with and without RA, showing that the DNA methylation loss observed at minor satellite sequences in *lsh*$^{-/-}$ 2i mESCs is not recovered during differentiation in the absence of *Lsh*. The hybridisation of the corresponding *MspI* digests with the same minor satellite probe acted as a control for digestibility and loading of the genomic DNA samples analysed (Figure 5.7 A and B, right panels). The probes for analysis of both major and minor satellites by Southern blotting were kindly provided by Professor Nick Gilbert.

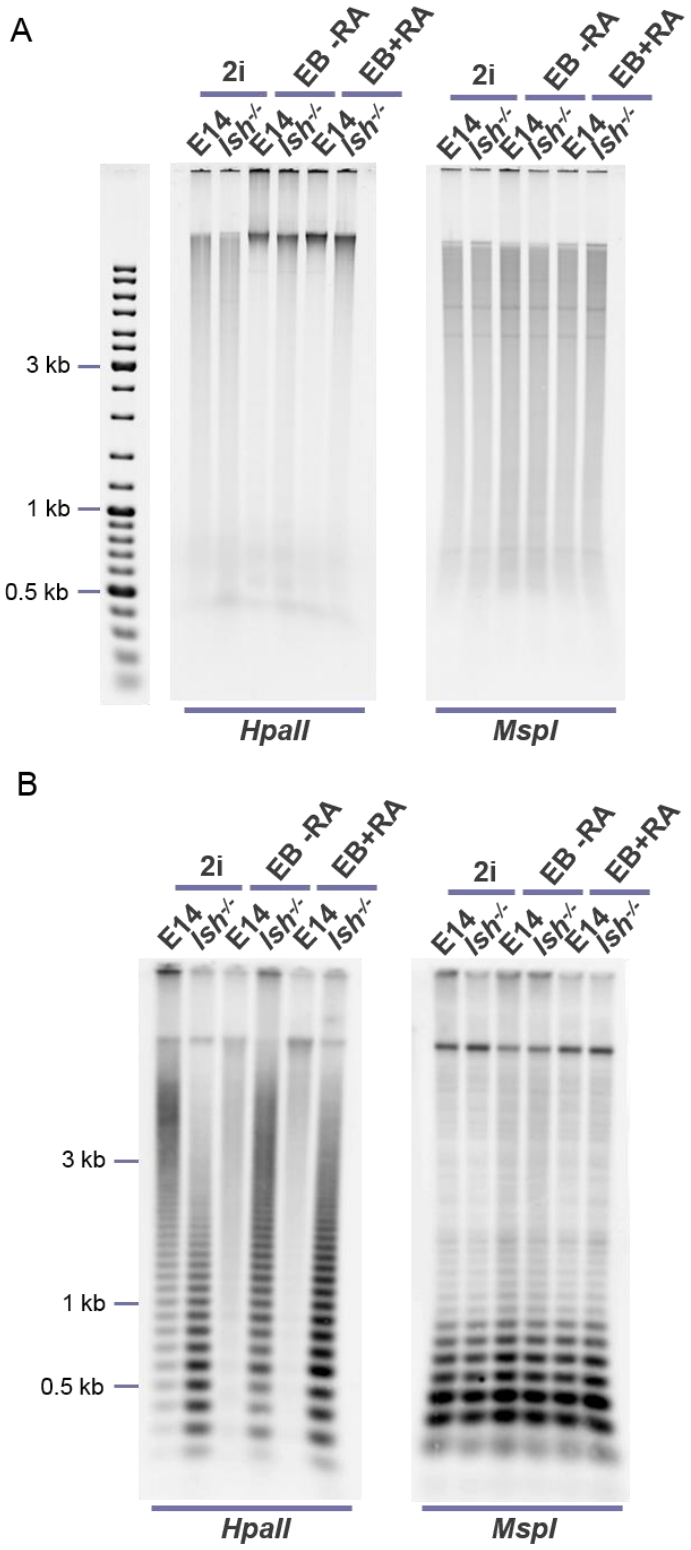


Figure 5.7. DNA methylation is incompletely established at minor satellites during differentiation of *Ish*$^{-/-}$ EBs in the presence and absence of RA. A. Agarose gels of E14 and *Ish*$^{-/-}$ 2i, EB –RA and EB +RA *HpaII* and *MspI* restriction digests prior to Southern blot transfer. B. Southern blot of *HpaII* and *MspI* digests, hybridised using a probe against minor satellites.

Southern blotting coupled with methylation-sensitive restriction digestion gave a strong indication of the relative levels of DNA methylation at major and minor satellites in E14 and *lsh*$^{-/-}$ 2i mESCs and EBs. This analysis was extended to quantify the DNA methylation levels at these repetitive regions using bisulfite sequencing, which assesses the 5-mC status of all CpGs within a specified region, rather than only those residing in specific restriction sites. Bisulfite-treated genomic DNA from E14 and *lsh*$^{-/-}$ mESCs and EBs was subjected to amplification using primers specific for major and minor satellites (primers described in Dunican *et al.*, 2013) to allow assessment and quantification of CpG methylation in these amplified sequences. Analysis of satellite DNA methylation in serum mESCs as well as 2i mESCs was included to further uncover the role of *Lsh* in facilitating methylation at these regions in different pluripotent states. This analysis revealed a relatively similar level of major satellite DNA methylation in E14 and *lsh*$^{-/-}$ serum mESCs of 79% and 71%, respectively (Figure 5.8 A). CpG methylation at major satellites was reduced equally to 51% and 55% upon adaptation to 2i in E14 and *lsh*$^{-/-}$ mESCs, respectively. Following differentiation to EBs, DNA methylation was re-established up to 75% in E14 EBs without RA and 80% in E14 EBs with RA, 5-mC levels which are comparable to those observed in serum E14 mESCs. However, this establishment of DNA methylation was impeded in *lsh*$^{-/-}$ EBs generated in the presence and absence of RA, which retained CpG methylation levels similar to those of *lsh*$^{-/-}$ 2i mESCs (55% in *lsh*$^{-/-}$ EBs without RA and 53% in *lsh*$^{-/-}$ EBs with RA). This indicates a crucial role for *Lsh* in *de novo* methylation of major satellites during differentiation.

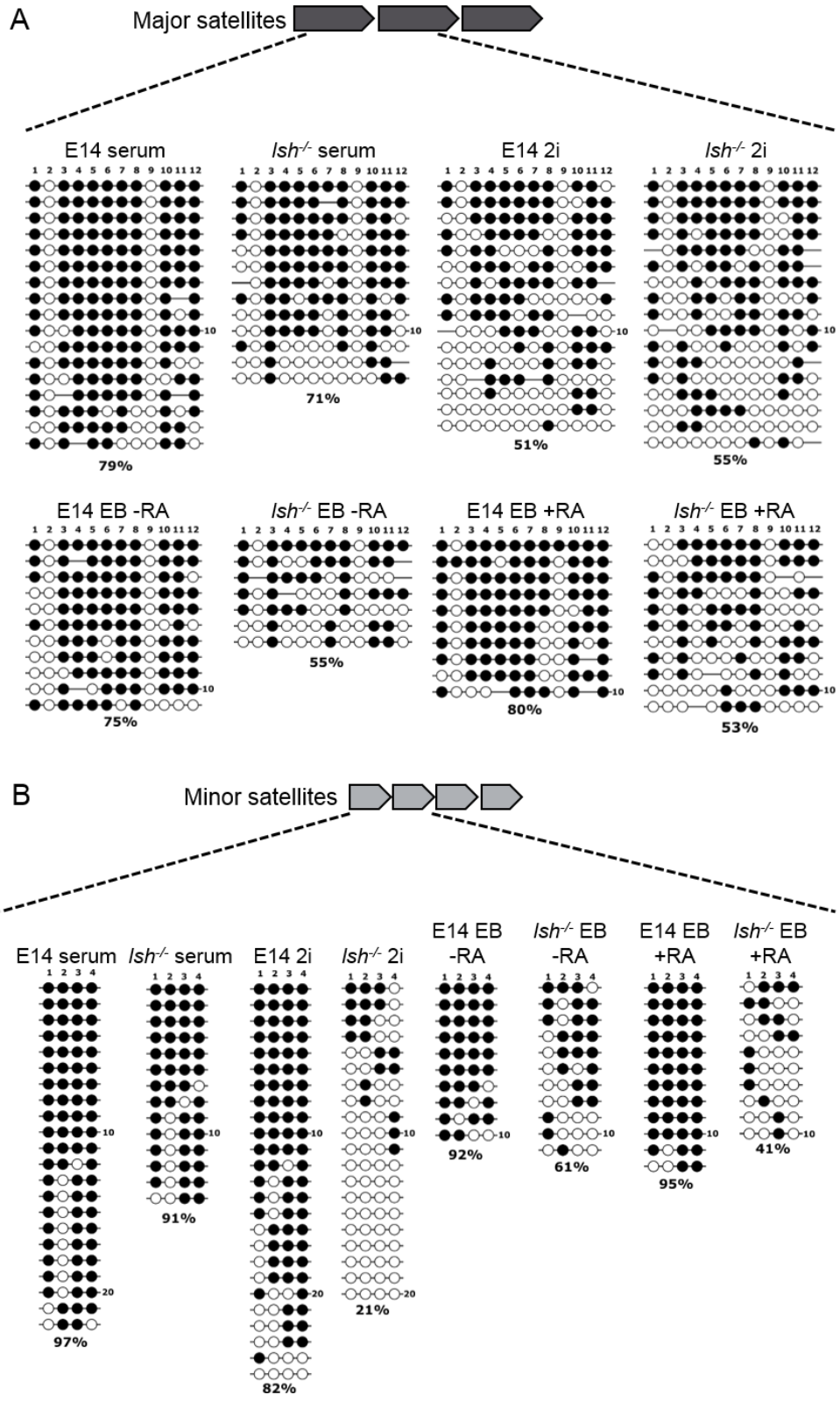


Figure 5.8. DNA methylation quantification at major and minor satellites in E14 and *lsh*$^{-/-}$ EBs. Quantification of major (A) and minor (B) satellite DNA methylation in E14 and *lsh*$^{-/-}$ EBs +/- RA by bisulfite sequencing. Filled circles represent methylated CpGs while unfilled circles are unmethylated CpGs. Number of CpGs assayed is shown at the top of each plot. Percentage of methylated CpGs is displayed under each plot.

At minor satellites, levels of 5-mC were comparable between E14 and *lsh*$^{-/-}$ serum mESCs, at 97% and 91% respectively (Figure 5.8 B). This level of CpG methylation was slightly reduced in E14 2i mESCs which showed 82% total DNA methylation. However, in accordance with the Southern blot shown in Figure 5.7 B, 5-mC at minor satellites was drastically reduced to just 21% in *lsh*$^{-/-}$ 2i mESCs. During differentiation to EBs, DNA methylation was re-established in E14 EBs cultured with and without RA, resulting 95% and 92% CpG methylation respectively, levels similar to those observed in E14 serum mESCs. In *lsh*$^{-/-}$ EBs with and without RA, partial DNA methylation re-establishment had occurred, although it only reached 41% and 61%, respectively. These results demonstrate that *Lsh* is required for full establishment of DNA methylation at minor satellites upon differentiation to EBs, while *Lsh* also appears to participate in maintenance of DNA methylation at minor satellites in the ground state of pluripotency.

There are limitations of the bisulfite sequencing technique that should be taken into consideration when interpreting results from bisulfite sequencing data. Incomplete bisulfite conversion can result in incorrect interpretation of unconverted unmethylated cytosines as methylated cytosines. Furthermore, the presence of clonal sequences, which are amplified from the same template molecule in the PCR, can skew the calculation of percentage methylation during analysis of bisulfite sequencing data. The use of the BISMA online tool (Rohde *et al.*, 2010), which has built-in algorithms for detection and removal of clonal sequences and sequences affected by incomplete conversion, reduces the impact of these limitations on the bisulfite sequencing analysis undertaken in this study. However, analysis of further clones for each sample is required to confirm the results and ensure their reproducibility.

Another limitation of the bisulfite sequencing technique that is more difficult to overcome is its inability to distinguish between 5-mC and 5-hmC. However, since the bisulfite sequencing results mirror those of the Southern blotting analysis in respect to differences in DNA methylation levels between E14 and *lsh*$^{-/-}$ 2i mESCs and EBs, it can be reasoned that these differences in DNA methylation levels are largely due changes in 5-mC rather than 5-hmC. Therefore, when taken together, the findings from the Southern blotting and bisulfite sequencing analysis uncover the critical but

differential role that *Lsh* plays in facilitating *de novo* DNA methylation at major and minor satellites during differentiation from naïve mESCs to EBs. Furthermore, these results highlight the additional requirement for *Lsh* in preserving the hypermethylated status of minor satellites in the pluripotent ground state.

5.2.2 Transition of 2i mESCs to EpiSCs

5.2.2.1 Experimental strategy

Analysis of *lsh*$^{-/-}$ EBs revealed a defect in *de novo* methylation of major and minor satellites and globally during differentiation from ground state mESCs. EBs are purported to be representative of the *in vivo* gastrulation process, due to the formation of all three germ layers, a primitive streak-like region and establishment of anteroposterior polarity (ten Berge *et al.*, 2008). Therefore, *Lsh* appears to contribute to the establishment of DNA methylation during differentiation from the naïve pluripotent state to these gastrula-like structures *in vitro*.

In Chapter 3, a culture system was used to examine the *de novo* DNA methylation that coincides with transition from 2i mESCs to serum mESCs, where 2i mESCs represented a naïve state of pluripotency and serum mESCs purportedly reflected a slightly later primed state that retains expression of naïve pluripotency markers, but also exhibits upregulated expression of lineage-associated factors. Investigation into the impact of absence of *Lsh* on *de novo* DNA methylation demonstrated that *Lsh* is not required for global establishment of DNA methylation during this transition. Therefore, I aimed to use a culture model that reflected an intermediate developmental state between serum mESCs and EBs to further deduce the exact developmental timepoint when *Lsh* is required to facilitate DNA methylation during the genome-wide establishment of 5-mC in the epiblast during implantation.

EpiSCs can be isolated from the post-implantation epiblast and maintained in culture as an *in vitro* model of a truly primed pluripotent state (Brons *et al.*, 2007; Tesar *et al.*, 2007; Nichols and Smith, 2009). EpiSCs can also be generated *in vitro* from mESCs continuously cultured in the presence of Activin A and FGF2 (Guo *et al.*, 2009). EpiSCs represent a developmentally and functionally distinct state from mESCs. They retain a degree of developmental potency as they have the capacity to differentiate into

multiple lineages *in vitro*, retain expression of core pluripotency factors *Oct4* and *Sox2* and are able to contribute to chimera formation when grafted into post-implantation embryos in culture (Guo *et al.*, 2009; Huang *et al.*, 2012; Kojima *et al.*, 2014). However, they are unable to efficiently contribute to chimera formation at the blastocyst stage - considered as a benchmark of functional pluripotency. EpiSCs are therefore proposed to reflect a more developmentally advanced state than mESCs, equivalent to the post-implantation epiblast (Huang *et al.*, 2012).

Similar to the post-implantation epiblast, EpiSCs are globally hypermethylated (Senner *et al.*, 2012; Habibi *et al.*, 2013; Veillard *et al.*, 2014). *In vitro* differentiation from mESCs to EpiSCs shares some epigenetic features with the equivalent process *in vivo*, such as DNA methylation establishment at one copy of the X chromosome in female cells (Guo *et al.*, 2009). Therefore, I aimed to explore the impact of *Lsh* deletion on the establishment of DNA methylation that occurs during this transition *in vitro*. To achieve this, hypomethylated E14 and *lsh*$^{-/-}$ mESCs adapted to 2i for 15 days (six passages) were cultured continuously in the presence of Activin A and FGF2 on a fibronectin-coated surface to differentiate them to EpiSCs (Figure 5.9). This involved maintaining cultures for multiple passages until robust, stably-growing EpiSC cultures had been established, usually taking 4-6 weeks. Morphological changes were assessed to monitor the transition of 2i mESCs to EpiSCs, as well as key transcriptional changes involving pluripotency and lineage-specific genes that are reported to change upon transition to a post-implantation developmentally restricted state *in vitro* and *in vivo* (experimental strategy illustrated in Figure 5.9).

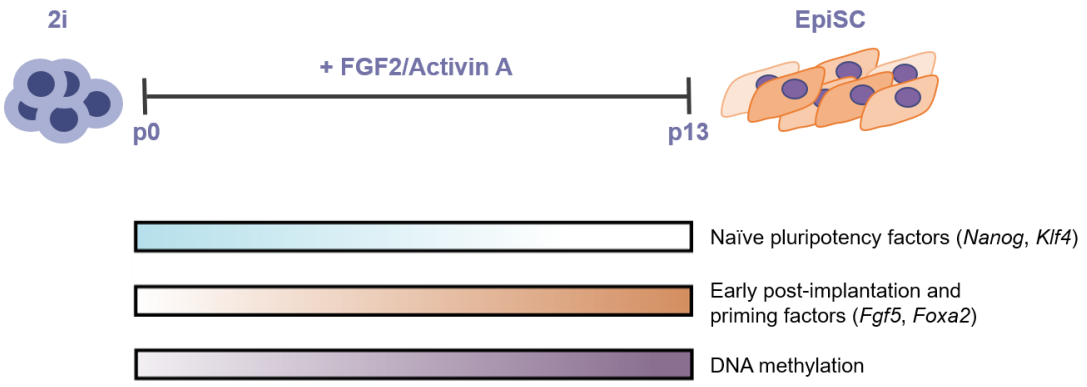


Figure 5.9. Schematic illustrating the transition of 2i mESCs to EpiSCs following continuous culture in FGF2 and Activin A. Key changes in expression of naïve pluripotency factors and early lineage specification markers are shown, as well as changes in DNA methylation.

5.2.2.2 E14 and lsh$^{-/-}$ 2i mESCs change morphologically during transition to EpiSCs

Mouse ESCs undergo extensive morphological changes during the transition to EpiSCs. The morphology of EpiSCs is similar to that of differentiated cells, as they exhibit growth in monolayers of flattened, irregularly-shaped cells (Guo *et al.*, 2009). At the start of the EpiSC transition, ground state E14 and *lsh*$^{-/-}$ mESCs exhibited typical 2i morphology as described previously, although the *lsh*$^{-/-}$ mESCs to a slightly lesser extent (Figure 5.10, left panels). Both E14 and *lsh*$^{-/-}$ mESCs displayed clear morphological changes throughout the transition to EpiSCs that resulted in a population of cells displaying the morphological characteristics of differentiated cells – flattened, irregularly-shaped cells that grow in a monolayer rather than in colonies (Figure 5.10, right panels). The populations of EpiSCs display a degree of morphological heterogeneity, particularly well-illustrated in the brightfield image of E14 EpiSCs in which a variety of cell morphologies are visible including large single cells, three-dimensional colonies of cells and cells with neuronal-like projections (indicated by the arrows in Figure 5.10). This morphological variation is consistent with previous reports that EpiSCs in culture represent a heterogeneous population of cells displaying a degree of transcriptional and functional variation (Han *et al.*, 2010; Kojima *et al.*, 2014; Tsakiridis *et al.*, 2014). Overall, both E14 and *lsh*$^{-/-}$ 2i mESCs display clear and widespread morphological changes during the transition to EpiSCs

indicative of adaptation to a culture state distinct from the naïve pluripotent ground state.

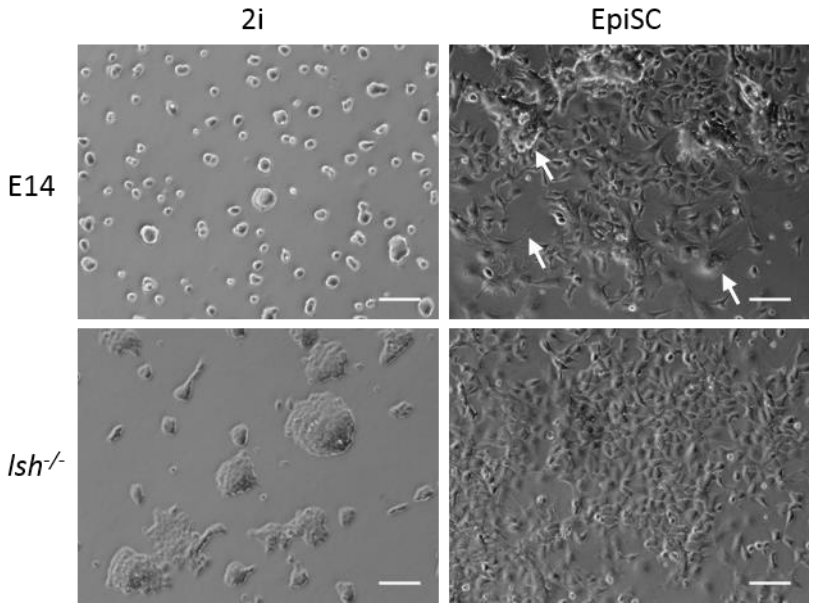


Figure 5.10. E14 and *lsh*$^{-/-}$ mESCs both exhibit morphological changes during transition to EpiSCs. Brightfield images of E14 and *lsh*$^{-/-}$ mESCs cultured in 2i (left panels) and after addition of FGF2/Activin A for 9 passages (right panels). Arrows highlight morphological features indicating a heterogeneous culture – three-dimensional colonies, neuronal-like projections and large single cells. Scale bars represent 100 µm.

5.2.2.3 E14 and lsh$^{-/-}$ mESCs exhibit mRNA expression changes indicative of transition to EpiSCs

The transcriptomes of naïve mESCs and EpiSCs are distinct. Although EpiSCs retain expression of core pluripotency factors *Oct4* and *Sox2*, they are devoid of expression of all naïve pluripotency factors, except for low levels of *Nanog* expression (Guo *et al.*, 2009; Kojima *et al.*, 2014). This loss of naïve pluripotency marker expression is accompanied by transcriptional upregulation of lineage-specific genes, resulting in an expression profile similar to that of the post-implantation epiblast (Guo *et al.*, 2009; Huang *et al.*, 2012; Kojima *et al.*, 2014). Therefore, expression of key pluripotency and lineage markers were evaluated during the transition from 2i mESCs to EpiSCs to assess whether conversion to a cell state reflective of post-implantation development had occurred. qRT-PCR expression analysis for two biological replicates (Rep 1 and Rep 2), each of which is an average of three technical replicates, is displayed in Figure 5.11.

The mRNA expression of key naïve pluripotency factors *Nanog* and *Klf4* was measured in E14 and *lsh*$^{-/-}$ 2i mESCs and EpiSCs by qRT-PCR. This demonstrated similarly relatively high expression of *Nanog* and *Klf4* in both E14 and *lsh*$^{-/-}$ mESCs adapted to 2i, although there are differences in the *Nanog* and *Klf4* mRNA levels between the two biological replicates. Upon transition to EpiSCs, *Nanog* expression was largely attenuated in both E14 and *lsh*$^{-/-}$ cells despite the initial expression level differences, although a slightly higher level of *Nanog* expression was retained in *lsh*$^{-/-}$ EpiSCs compared to E14 in the first biological replicate (Figure 5.11 A). Concurrently, mRNA levels of *Klf4* were reduced to an almost undetectable level in both biological replicates of E14 and *lsh*$^{-/-}$ EpiSCs. Transcriptional downregulation of these two key naïve pluripotency markers indicate an exit from ground state pluripotency, suggesting that both E14 and *lsh*$^{-/-}$ mESCs are transitioning to a more developmentally restricted cellular state.

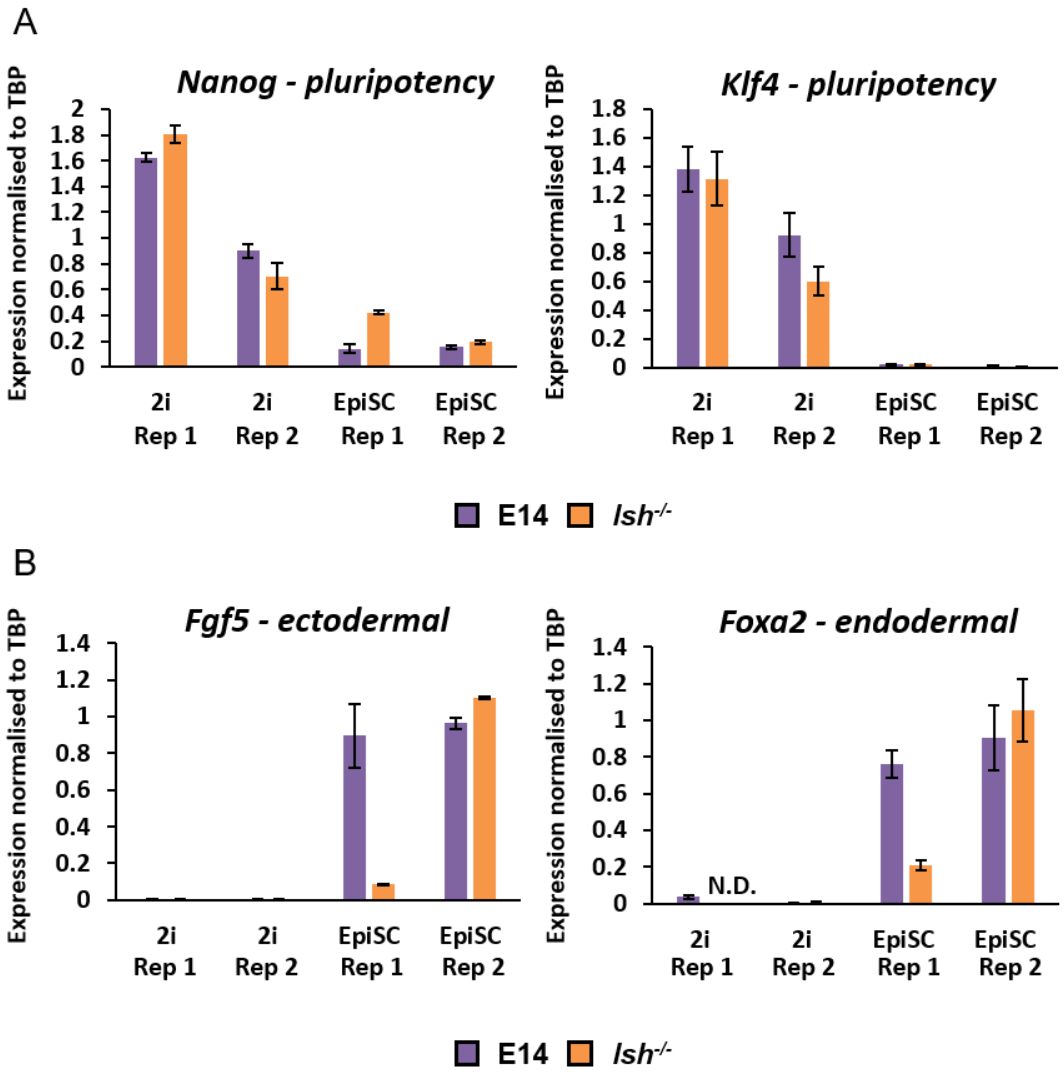


Figure 5.11. E14 and *lsh*$^{-/-}$ cells display transcriptional changes in key genes following transition to EpiSCs. Assessment of mRNA expression by qRT-PCR of key pluripotency genes (A) and lineage-associated factors (B) in E14 and *lsh*$^{-/-}$ cells before and after transition to EpiSCs. Results from two biological replicates are shown independently for each cell line and culture condition and are labelled as Rep 1 and Rep 2. All genes are normalised to *TBP* expression. Error bars represent +/- propagated standard deviation of three technical replicates. N.D. = not detected.

Transcriptional silencing of naïve pluripotency factors upon transition of mESCs to EpiSCs coincides with upregulation of lineage associated genes, such as *Fgf5*, an ectodermal marker expressed in the early post-implantation epiblast, and the early endodermal marker *Foxa2* (Tesar *et al.*, 2007; Guo *et al.*, 2009; Kojima *et al.*, 2014). In 2i E14 and *lsh*$^{-/-}$ mESCs, levels of *Fgf5* and *Foxa2* mRNA are extremely low in both biological replicates (Figure 5.11 B). During transition to EpiSCs, expression of these lineage-associated factors is massively upregulated in E14 EpiSCs. *Fgf5* and *Foxa2*

mRNA levels were upregulated to a similar degree in the second biological replicate for *lsh*$^{-/-}$ EpiSCs. However, although *Fgf5* and *Foxa2* expression is hugely increased in the first replicate of *lsh*$^{-/-}$ EpiSCs compared to the corresponding 2i mESCs, they are not upregulated to the same extent as in E14 EpiSCs or the other *lsh*$^{-/-}$ EpiSC biological replicate. Indeed, for the first biological replicate, there is a 10-fold difference in *Fgf5* mRNA levels between E14 and *lsh*$^{-/-}$ EpiSCs and a 3.5-fold difference in *Foxa2* expression (Figure 5.11 B, Rep 1). Coupled with the incomplete repression of *Nanog* in the first biological replicate for *lsh*$^{-/-}$ EpiSCs observed in Figure 5.11 A, this suggests that the transition to a transcriptional state reflective of the post-implantation epiblast was impaired in the first biological replicate.

Despite the much reduced scale of *Fgf5* and *Foxa2* upregulation in *lsh*$^{-/-}$ EpiSCs compared to E14 in the first biological replicate, the general trend of expression changes in these factors is the same between the E14 and *lsh*$^{-/-}$ across both biological replicates. Taken together with the downregulation of *Nanog* and *Klf4* in E14 and *lsh*$^{-/-}$ EpiSCs, and the apparent morphological changes exhibited by both cell lines, there appears to be a degree of transition to an EpiSC-like state in both E14 and *lsh*$^{-/-}$ cells for both biological replicates.

5.2.2.4 Global DNA methylation levels are reduced in lsh$^{-/-}$ 2i mESCs and EpiSCs compared to E14

To examine the impact of the absence of *Lsh* on global DNA methylation establishment during transition of 2i mESCs to EpiSCs, total 5-mC levels were measured by LC-MS in E14 and *lsh*$^{-/-}$ 2i mESCs and EpiSCs. The average percentage methylation of the two biological replicates shown in Figure 5.11 is presented in Figure 5.12. Consistent with the global DNA methylation LC-MS data shown in Figure 5.5, a significant difference in 5-mC levels of approximately 22% between E14 and *lsh*$^{-/-}$ 2i mESCs is evident (Figure 5.12). Following transition to EpiSCs, there is also a significant difference of 18% between E14 and *lsh*$^{-/-}$ cultures. This demonstrates impaired establishment or maintenance of global DNA methylation in the absence of *Lsh* in 2i mESCs and EpiSCs, implying that *Lsh* is required for wild-type global DNA methylation levels in these culture states. Furthermore, the size of the error bars apparent in Figure 5.12, representing the standard error of the mean of the two

biological replicates, indicates that the global DNA methylation levels are very similar between biological replicates despite the differences in expression of key pluripotency and lineage-associated factors shown in Figure 5.11. This suggests that the changes in global DNA methylation levels that occur during transition of 2i *lsh*$^{-/-}$ mESCs to EpiSCs occur independently of the changes in expression of key pluripotency and lineage-associated genes.

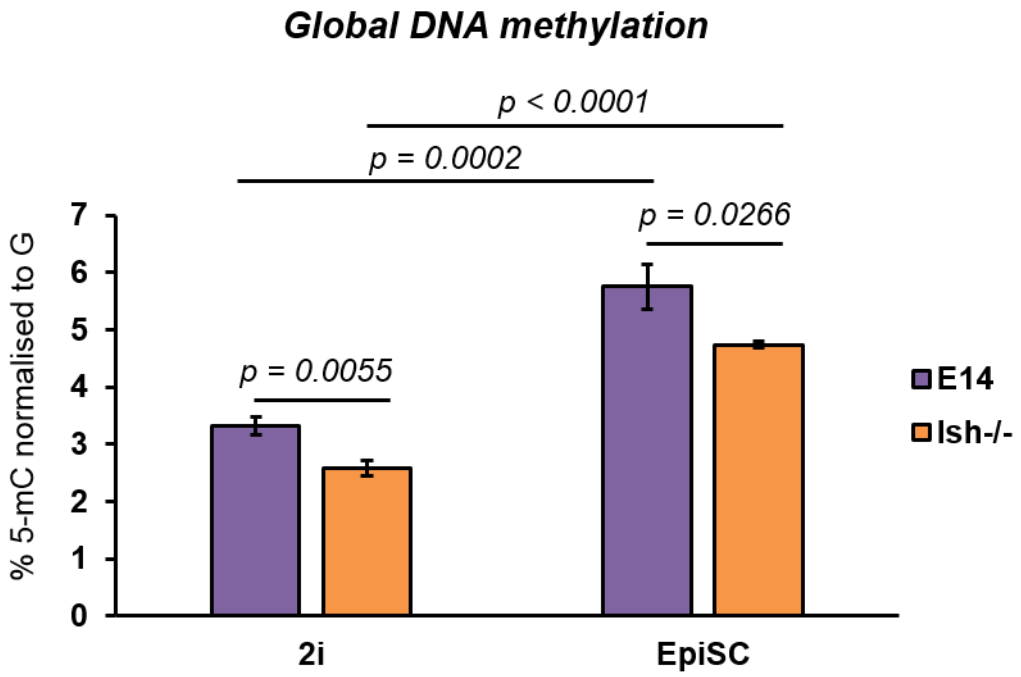


Figure 5.12. Global DNA methylation levels are not fully established in *lsh*$^{-/-}$ EpiSCs. Measurement of global 5-mC levels by LC-MS in E14 and *lsh*$^{-/-}$ 2i mESCs and EpiSCs. Error bars represent +/- standard error of two biological replicates. P-values indicate level of significance between 2i and EpiSC E14 and *lsh*$^{-/-}$ cells as measured by a two-way unpaired t-test.

5.2.2.5 DNA methylation establishment at major and minor satellites is impaired in the absence of Lsh during transition to EpiSCs

To investigate whether a similar defect in *de novo* methylation of major and minor satellites was apparent during differentiation to of *lsh*$^{-/-}$ 2i mESCs to EpiSCs, I used the methylation-sensitive *MaeII* and *HpaII* digests combined with Southern blotting as described previously, using DNA from the first biological replicate shown in Figure 5.11. *MaeII* digestion of genomic DNA from E14 and *lsh*$^{-/-}$ 2i mESCs revealed similar levels of DNA hypomethylation in these samples, concurrent with the same digest

shown in Figure 5.6 A (Figure 5.13 A). However, a dramatic reduction in DNA methylation at *MaeII* sites was apparent in *lsh*$^{-/-}$ EpiSCs compared to E14, demonstrated by the increased presence and intensity of lower molecular weight bands (indicated by the arrows in Figure 5.13 A). Digestion of genomic DNA from these cells with *HpaII*, which is generally indicative of global DNA methylation, agreed with marked DNA hypomethylation of *lsh*$^{-/-}$ EpiSCs observed in the *MaeII* digest (Figure 5.13 A). Furthermore, the *HpaII* digest corroborated the LC-MS results, indicating that global 5-mC levels are slightly lower in *lsh*$^{-/-}$ 2i mESCs compared to E14.

To provide a more comprehensive estimation of 5-mC levels at major and minor satellites in E14 and *lsh*$^{-/-}$ EpiSCs, the *MaeII* and *HpaII* digests were again coupled with Southern blotting using probes for these repetitive sequences. Hybridisation of *MaeII* digests using a probe for major satellites strengthened the observation that while these regions become hypermethylated in E14 EpiSCs, DNA methylation establishment at major satellites is severely impaired in *lsh*$^{-/-}$ EpiSCs (Figure 5.13 B). Analysis of the 5-mC status of minor satellites using a probe for these sequences on the *HpaII* digests revealed a similar pattern of DNA methylation of these regions as shown previously in Figure 5.7 B. Again, a striking hypomethylation of minor satellites in *lsh*$^{-/-}$ 2i mESCs was evident (Figure 5.13 C). This reduction in DNA methylation at minor satellites in the absence of *Lsh* was not recovered following transition to EpiSCs, which also showed substantial hypomethylation in *lsh*$^{-/-}$ cells compared to hypermethylated E14 EpiSCs.

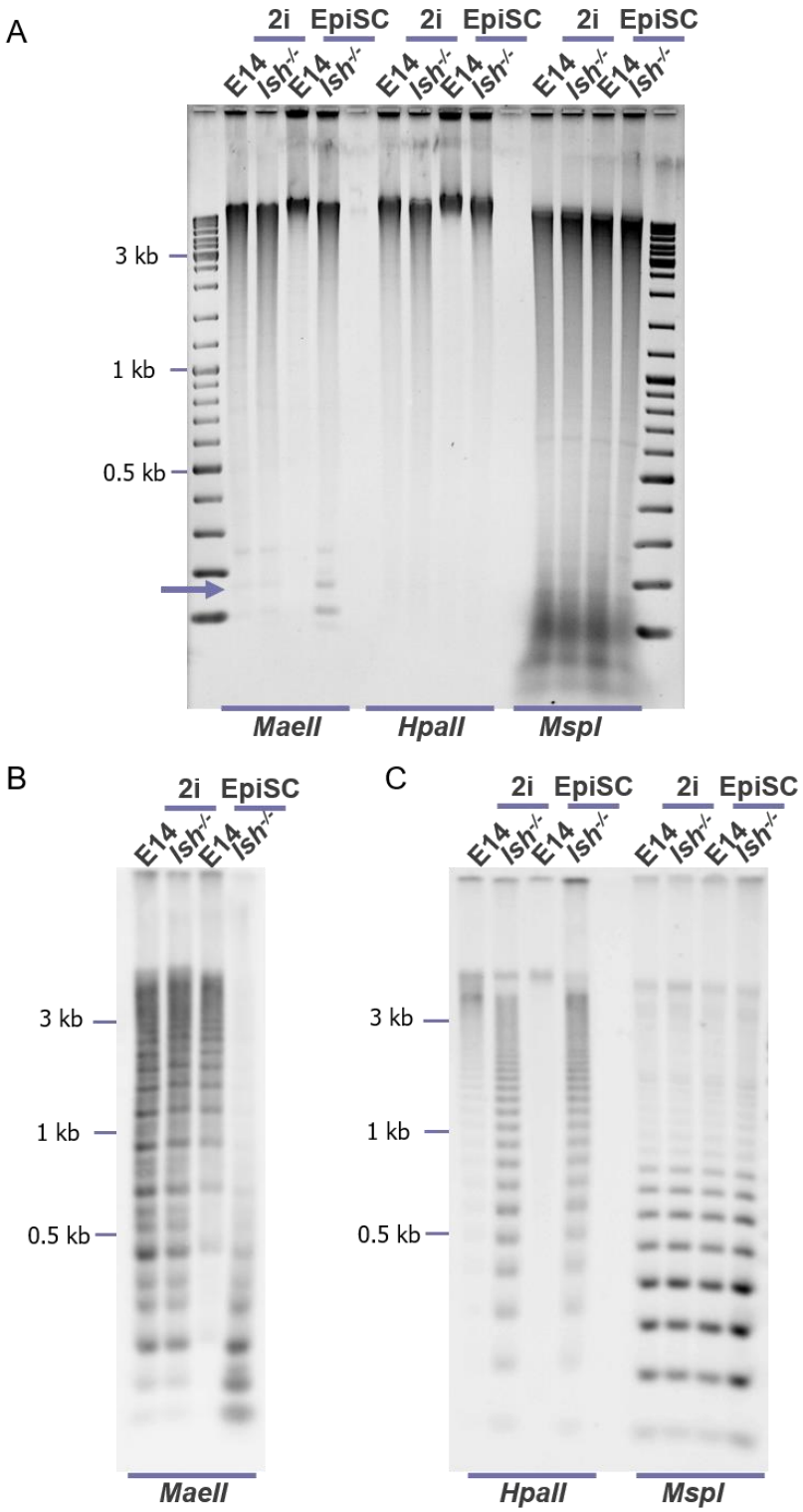


Figure 5.13. Major and minor satellites are hypomethylated in *Ish*$^{-/-}$ EpiSCs. A. Agarose gel displaying DNA from E14 and *Ish*$^{-/-}$ 2i mESCs and EpiSCs restriction digested by *MaeII*, *HpaII* or *MspI* restriction enzyme, as indicated. B. Southern blot of *MaeII* digests displayed on agarose gel in (A) using probe against major satellite sequences. C. Southern blot of *HpaII* and *MspI* digests shown on gel in (A) using probe against minor satellite sequences.

The CpG methylation status of major and minor satellite repeats in E14 and *lsh*$^{-/-}$ EpiSCs was quantified at base pair resolution using bisulfite sequencing. The bisulfite sequencing plots from serum and 2i mESCs originally shown in Figure 5.8 are repeated here for comparative purposes. E14 EpiSCs exhibited a CpG methylation level of 76% at major satellites, comparable to those observed in serum E14 mESCs (Figure 5.14 A). However, as suggested by the Southern blotting analysis, establishment of DNA methylation was severely impeded in *lsh*$^{-/-}$ EpiSCs, which exhibited only 24% CpG methylation at major satellites. This is far lower than the 55% CpG methylation apparent in *lsh*$^{-/-}$ 2i mESCs, indicating that not only has DNA methylation not been established at major satellites during differentiation to EpiSCs, it was actually lost at these regions in the absence of *Lsh*. This implies a role for *Lsh* beyond that of promoting *de novo* methylation at pericentromeric repeats during post-implantation development, suggesting that it may also contribute to maintenance of methylation at these regions.

At minor satellites, E14 EpiSCs displayed very high 5-mC levels of 93%, similar to those of serum E14 mESCs and slightly higher than 2i E14 mESCs (Figure 5.14 B). In contrast, *lsh*$^{-/-}$ EpiSCs showed a markedly reduced level of DNA methylation of 56%. This was higher than the hypomethylated minor satellite sequences in *lsh*$^{-/-}$ 2i mESCs, which exhibit only 21% DNA methylation, suggesting that partial DNA methylation establishment had occurred during differentiation to EpiSCs in the absence of *Lsh*. Therefore, these results suggest that although partial establishment of DNA methylation occurs during differentiation from 2i mESCs to EpiSCs in the absence of *Lsh*, full DNA methylation establishment requires the presence of *Lsh*. These bisulfite sequencing results from EpiSCs mirror the findings from EBs, reinforcing the proposal that *Lsh* is required for full establishment of DNA methylation at major and minor satellites during differentiation from naïve pluripotency to a more developmentally restricted state.

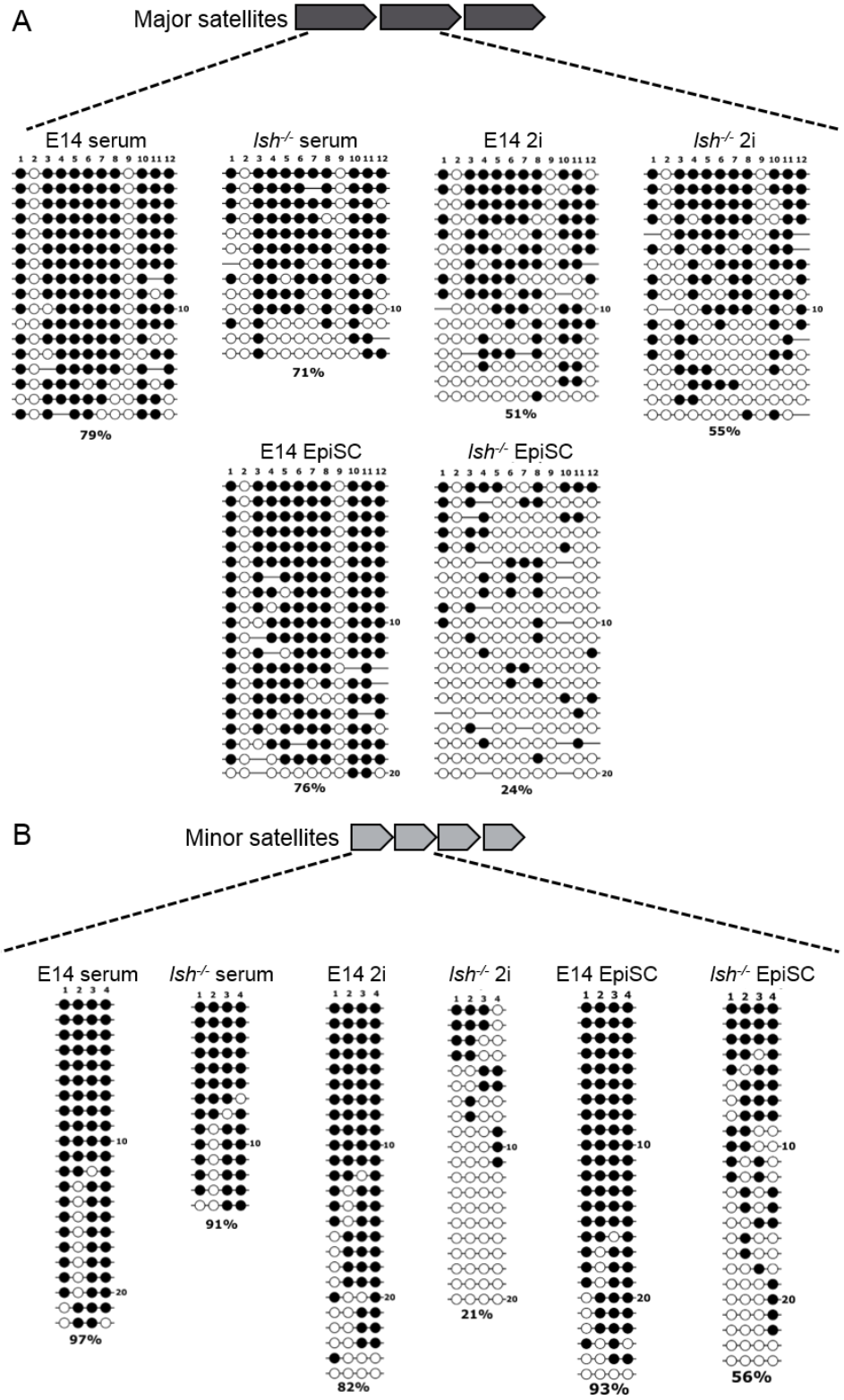


Figure 5.14. *Lsh* is required for WT levels of DNA methylation at major and minor satellites in EpiSCs. Quantification of major (A) and minor (B) satellite DNA methylation in E14 and *lsh*$^{-/-}$ serum and 2i mESCs and EpiSCs by bisulfite sequencing. Filled circles represent methylated CpGs while unfilled circles are unmethylated CpGs. Number of CpGs assayed is shown at the top of each plot. Percentage of methylated CpGs is displayed under each plot.

Collectively, these DNA methylation analyses provide further evidence for the crucial role that *Lsh* plays in facilitating DNA methylation at major and minor satellites and globally during the establishment of DNA methylation that occurs during differentiation. Furthermore, these results suggest that this critical function of *Lsh* is required during the transition to EpiSCs, prior to the gastrulation stage of development that EBs represent. This implies that *Lsh* contributes to the initial wave of *de novo* DNA methylation that occurs largely in the implanting blastocyst and the early post-implantation epiblast during embryonic development.

5.2.3 Transcriptome profiling of E14 and lsh$^{-/-}$ EpiSCs

Since *lsh*$^{-/-}$ EpiSCs exhibit DNA hypomethylation globally and at specific repeat sequences, I sought to investigate whether absence of *Lsh* also has an effect on gene expression patterns in EpiSCs. Therefore, gene expression profiles of E14 and *lsh*$^{-/-}$ EpiSCs were generated using RNA-seq from three biological replicates for each cell line, with bioinformatic analysis of the RNA-seq datasets performed by Dr Donncha Dunican.

5.2.3.1 The transcriptional profiles of EpiSCs are distinct from mESCs

To examine the degree of variation between the gene expression profiles of E14 and *lsh*$^{-/-}$ EpiSCs and the mESCs from which they were generated, PCA was undertaken. The transcriptional profiles of three biological replicates for E14 and *lsh*$^{-/-}$ EpiSCs were compared to the single RNA-seq replicate generated for serum and 2i E14 and *lsh*$^{-/-}$ mESCs, which were analysed in depth in Chapter 4. This revealed that the principal component of variation (85%) between the transcriptional profiles of mESCs cultured in serum or 2i and EpiSCs, highlighting the extensive remodelling of gene expression patterns that occurs during transition from ground state mESCs to EpiSCs (Figure 5.15). Furthermore, it demonstrates that transcriptional profiles cluster based on cellular state rather than the genotype, as shown for the PCA of gene expression profiles in Chapter 4. The secondary component of variation is accounted for by the differences between the transcriptomes of serum and 2i mESCs, indicating that the gene expression profiles of E14 and *lsh*$^{-/-}$ EpiSCs are somewhat similar.

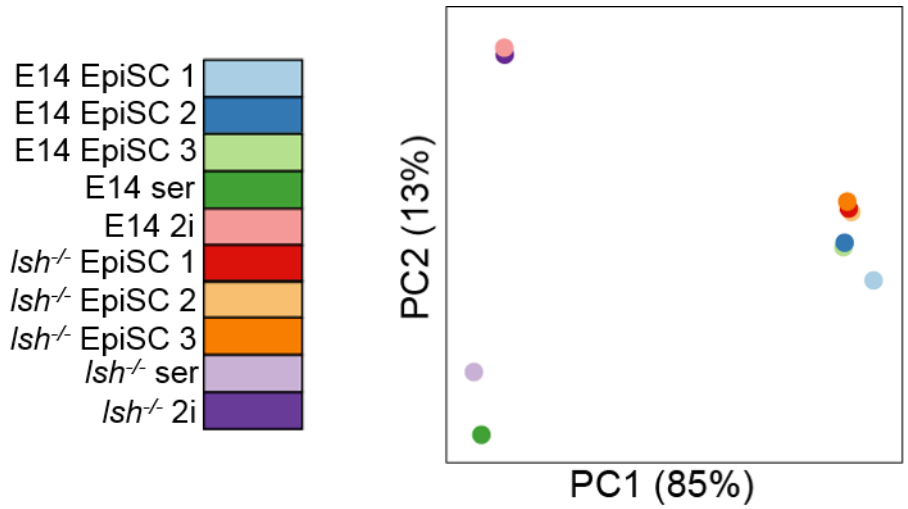


Figure 5.15. The transcriptional profiles of E14 and *lsh*$^{-/-}$ mESCs and EpiSCs principally cluster based on cellular state. PCA of gene expression profiles for serum and 2i E14 and *lsh*$^{-/-}$ mESCs and EpiSCs. Each replicate for EpiSCs is plotted individually.

5.2.3.2 Genes involved in developmental processes are mis-expressed in lsh$^{-/-}$ EpiSCs

To further investigate whether there are any differences in gene expression in EpiSCs lacking *Lsh*, the transcriptional profiles of *lsh*$^{-/-}$ EpiSCs were plotted against E14 EpiSCs and presented in a scatter plot. This uncovered over 1000 genes upregulated more than two-fold in *lsh*$^{-/-}$ EpiSCs and over 900 genes downregulated more than two-fold (Figure 5.16 A). GO term analysis of the upregulated genes demonstrated a significant enrichment in terms associated with developmental processes (Figure 5.16 B). Interestingly, there are also a number of terms linked with neuron development, as well as an enrichment in upregulated genes involved in behaviour. Many terms that are enriched in the GO analysis of the upregulated genes are also found in the GO analysis of the downregulated genes, such as “cell differentiation” and “multicellular organismal process” (Figure 5.16 C). GO terms associated with development dominate the enrichment analysis of the downregulated genes, alongside terms closely linked to developmental processes such as “anatomical structure morphogenesis”. These results reveal that genes involved in the regulation of cell and organism development are mis-expressed in the absence of *Lsh* in the context of EpiSCs. However, at this level of analysis, the mis-expression patterns do not appear to interfere with EpiSC identity.

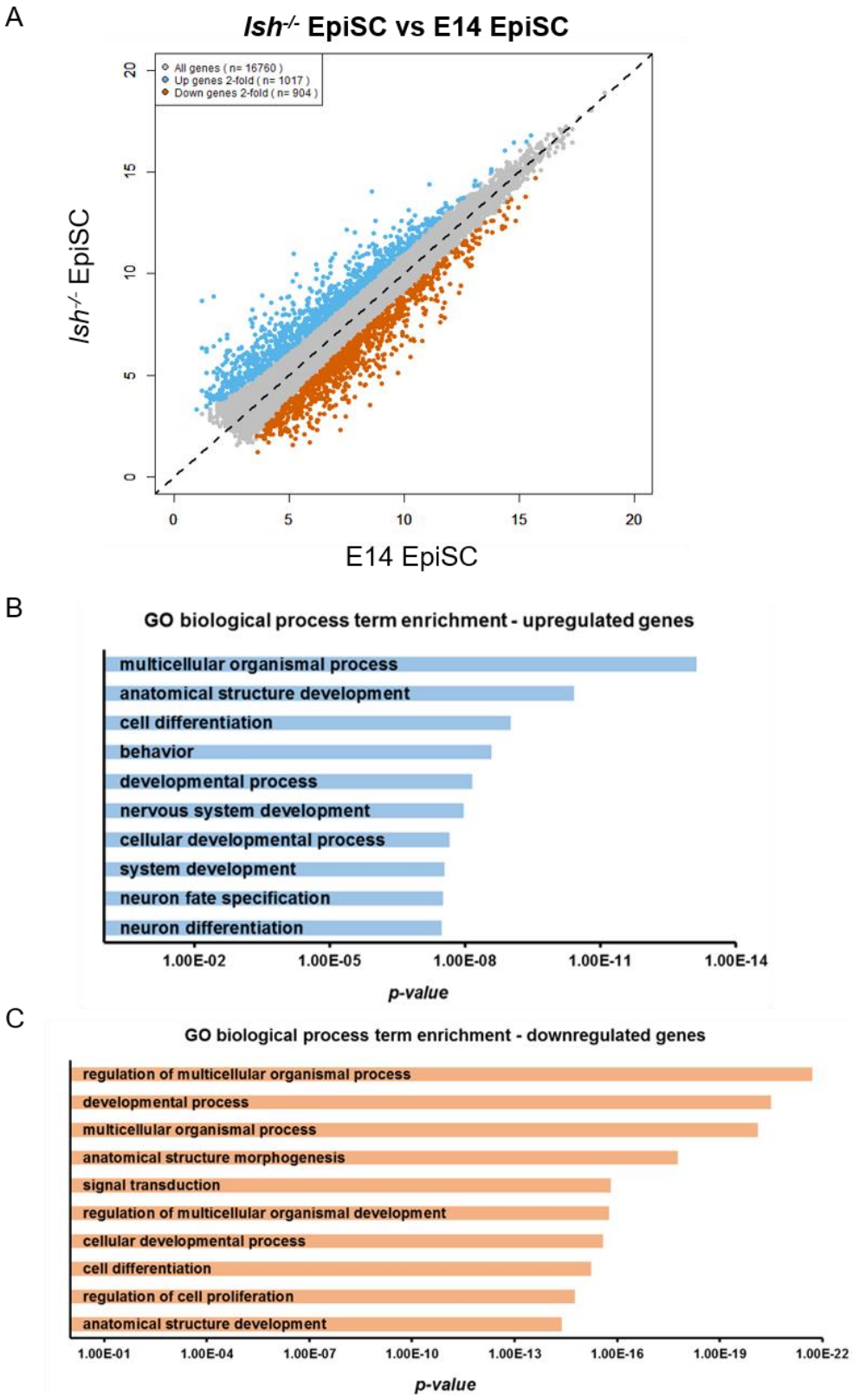


Figure 5.16. *Lsh* influences expression of genes involved in developmental processes in EpiSCs. A. Scatter plot comparing transcriptional profiles of E14 and *lsh*$^{-/-}$ EpiSCs. Genes upregulated more than two-fold are highlighted in blue, genes downregulated more than two-fold are shown in orange. Axes represent $\log_2$(read counts

per 10 million) for the sample indicated on each axis. Genes with read counts below 5 were excluded. B. GO term enrichment analysis of genes upregulated more than two-fold in *lsh*$^{-/-}$ EpiSCs. C. GO term enrichment analysis of genes downregulated more than two-fold in *lsh*$^{-/-}$ EpiSCs. GO analysis generated using the biological processes filter of the GOrilla analysis tool. Top ten significant GO terms are shown in both B and C.

5.2.4 Summary of DNA methylation perturbations in the absence of Lsh

Investigation into the impact of *Lsh* deletion on global and local DNA methylation levels across a variety of culture models has revealed specific requirements for *Lsh* in facilitating DNA methylation that are dependent on developmental context. The DNA methylation perturbations in the absence of *Lsh* across the culture models utilised in this study are summarised in Table 5.1. It highlights the specific requirement for *Lsh* in ground state mESCs to facilitate DNA methylation of minor satellites, while cytosine methylation of major satellites is most dependent on *Lsh* in EpiSCs. Overall, *Lsh* generally appears to be required to establish WT levels of DNA methylation at both major and minor satellites, as well as globally, following differentiation from ground state pluripotency. Additionally, I have explored the effect of mutation of human homologue *HELLS* in a somatic cell context and demonstrated that this results in global DNA hypomethylation (presented in Appendix Figure A2). This indicates that, as anticipated, the mouse and human homologues of *Lsh* share a common function in facilitating DNA methylation, and that the roles of *Lsh* in DNA methylation during mouse development could help to infer a similar function for *HELLS* in a human developmental context.

	Genomic region exhibiting DNA hypomethylation in an *lsh*$^{-/-}$ context		
Culture model	***Global***	***Major satellites***	***Minor satellites***
Serum mESCs	No	No	No
2i mESCs	Yes	No	Yes - severe
Reversion mESCs	No	No*	No*
EB -RA	Yes	Yes – moderate	Yes - moderate
EB +RA	Yes	Yes – moderate	Yes - moderate
EpiSC	Yes	Yes – severe	Yes - moderate
Human somatic cancer cell line	Yes**	-	-

Table 5.1. Table summarising DNA hypomethylated genomic regions in *lsh*$^{-/-}$ cells compared to the corresponding E14 sample.
* = data that has not been presented in this thesis
** = data shown in Appendix Figure A2

5.2.5 Lsh is transcriptionally downregulated during differentiation

Lsh is ubiquitously expressed at low levels throughout multiple tissues in adult mice, but exhibits specific enrichment in tissues with high proliferation and differentiation capacities, such as testis, thymus and bone marrow (Raabe *et al.*, 2001). However, *Lsh* is more broadly expressed at high levels in multiple mouse embryonic tissues, suggesting a more general role for *Lsh* throughout the developing embryo (Geiman *et al.*, 2001; Raabe *et al.*, 2001). I sought to determine *Lsh* expression across the culture models of pluripotency and post-implantation development that I had utilised in this study to determine whether the apparent contribution of *Lsh* to DNA methylation establishment coincided with transcriptional upregulation of *Lsh*. qRT-PCR analysis of three technical replicates from one biological replicate for each culture condition is presented in Figure 5.17. In fact, *Lsh* expression appears to decrease following differentiation to EpiSCs and EBs, where its absence has the most notable effect on global and satellite DNA methylation levels (Figure 5.17). *Lsh* expression is higher in the pluripotent culture states, particularly so in 'primed' mESCs cultured in serum, where its mRNA levels are almost two-fold higher than in naïve 2i mESCs. The potential reasons and significance for the apparent enrichment of *Lsh* in pluripotent cells is discussed in Section 5.3.6.

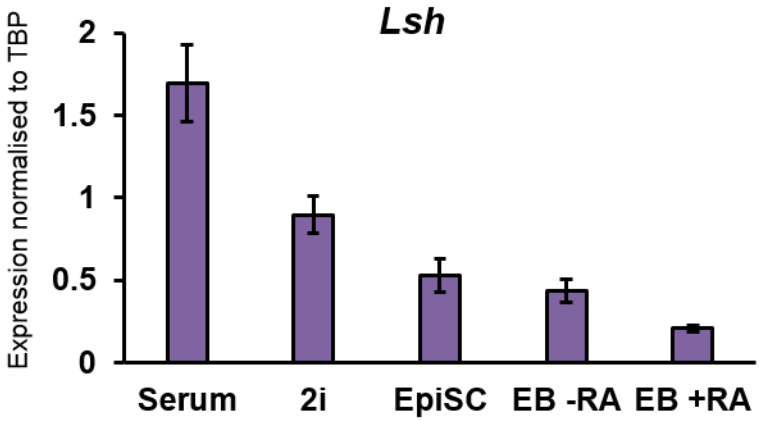


Figure 5.17. *Lsh* is transcriptionally downregulated upon differentiation. Assessment of mRNA expression by qRT-PCR of *Lsh* in E14 mESCs cultured in serum or 2i, and after differentiation to EpiSCs or EBs. All genes normalised to TBP expression. Error bars represent +/- propagated standard deviation of three technical replicates.

5.3 Discussion

In this chapter, I aimed to dissect the requirement for *Lsh* in facilitating DNA methylation during differentiation. To achieve this, I transitioned E14 and *lsh*$^{-/-}$ 2i mESCs to EBs and EpiSCs to examine the role of *Lsh* in the establishment of DNA methylation that accompanies the transition from the hypomethylated ground state to these hypermethylated, more developmentally restricted culture states. This revealed an apparent requirement for *Lsh* in establishing DNA methylation at major and minor satellites during differentiation. Previous reports into the function of *Lsh* in *de novo* DNA methylation during differentiation focused on the establishment of 5-mC at specific genomic loci in EBs generated in the presence of RA (Xi *et al.*, 2009; Ren *et al.*, 2015). The work presented in this chapter extended this analysis to examine global DNA methylation levels as well as specific repetitive regions. I also assessed establishment of DNA methylation in EBs generated in the absence of RA, and importantly, in EpiSCs following differentiation from 2i mESCs. This acted to narrow down the developmental window where the function of *Lsh* in influencing DNA methylation levels and patterns becomes relevant. Furthermore, exploration of DNA methylation patterns at centromeric and pericentromeric repeats uncovered differential regulation of DNA methylation by *Lsh* at these regions, with the genomic regions

affected by DNA hypomethylation in the absence of *Lsh* varying depending on cellular context.

5.3.1 Global and local DNA methylation defects following differentiation to EBs and EpiSCs

Analysis of global 5-mC levels in E14 and *lsh*$^{-/-}$ EBs and EpiSCs that had been differentiated from 2i mESCs revealed a significant reduction of 18-20% in global DNA methylation levels in the absence of *Lsh*. This implies that the establishment of *de novo* DNA methylation is impeded in absence of *Lsh* during differentiation to EBs and EpiSCs. Interestingly, LC-MS analysis also revealed that *lsh*$^{-/-}$ 2i mESCs exhibited significantly reduced global 5-mC levels (16-22%) prior to differentiation. This points to a previously unappreciated role for *Lsh* in regulating global DNA methylation levels in naïve pluripotency. However, it also makes it difficult to deduce whether the DNA hypomethylation observed in *lsh*$^{-/-}$ EpiSCs and EBs is due to a defect in *de novo* DNA methylation, or due to a pre-existing reduction in DNA methylation levels. Southern blotting and bisulfite sequencing analysis of DNA methylation levels at major and minor satellites provided further evidence to assist the interpretation of these global DNA methylation results. At major satellites, no DNA hypomethylation was apparent in *lsh*$^{-/-}$ 2i mESCs compared to E14, yet upon differentiation to EBs DNA methylation levels increased in E14 EBs, whereas this increase did not occur in *lsh*$^{-/-}$ EBs. Since there was no DNA hypomethylation at major satellites prior to differentiation in *lsh*$^{-/-}$ 2i mESCs, this suggests that, at least at these genomic regions, attenuation of *de novo* methylation is the causative factor resulting in reduced 5-mC levels at major satellites in EBs. This implies that, at least at some genomic regions, the global hypomethylation evident in the absence of *Lsh* following differentiation cannot be accounted for by pre-existing DNA hypomethylation in *lsh*$^{-/-}$ 2i mESCs.

The significant difference in global 5-mC levels between E14 and *lsh*$^{-/-}$ 2i mESCs observable by LC-MS in this chapter was not evident in the LC-MS data presented in Chapter 3 (Figure 3.9). This inconsistency between the LC-MS datasets could be explained by a number of technical and biological factors. For the 5-mC measurements displayed in this chapter, the LC-MS protocol, spectrometry and data analysis methods had been improved, increasing the specificity and ability to detect more nuanced

differences in 5-mC content. This could account for the difference that is now apparent in *lsh*$^{-/-}$ 2i mESCs that was not evident previously.

The difference in 5-mC content between E14 and *lsh*$^{-/-}$ 2i mESCs that was not observed initially could also be explained by the extended period of culture that the mESCs in this chapter had undergone compared to the mESCs used in earlier analysis. This may have resulted in a progressive loss in DNA methylation due to impaired maintenance methylation since the deletion of *Lsh*, a phenomenon that has been described for *Dnmt3a/3b*$^{-/-}$ double null mESCs (Jackson *et al.*, 2004). This highlights a potential role for *Lsh* in maintenance of DNA methylation globally in the ground state of pluripotency, a concept that has been supported by the analysis of DNA methylation at major and minor satellites presented in this chapter.

5.3.2 Lsh contributes to de novo DNA methylation during transition to EpiSCs

The main aim of work presented in this chapter was to narrow down the developmental window where *Lsh* function becomes relevant for influencing DNA methylation establishment during development. Since the absence of *Lsh* did not affect the establishment of DNA methylation during transition from hypomethylated ground state 2i mESCs to hypermethylated primed serum mESCs, I transitioned 2i mESCs to culture states representing later developmental time points. First, I differentiated 2i mESCs to EBs, which are proposed to represent the gastrulation stage of embryonic development where establishment of DNA methylation is purported to influence and reinforce cell fate decisions during germ layer specification. Here, I demonstrated that *Lsh* was required to establish WT levels of DNA methylation globally and at satellite repeats in EBs generated in the presence and absence of RA. This indicated that *Lsh* facilitates establishment of DNA methylation between the naïve state of pluripotency and the gastrula stage of development. I then transitioned 2i mESCs to EpiSCs to determine whether this failure to establish DNA methylation was apparent at the earlier post-implantation stage of development that EpiSCs are purported to reflect. Indeed, a similar degree of global and local satellite hypomethylation was evident in *lsh*$^{-/-}$ mESCs, suggesting that *Lsh* primarily acts during this earlier developmental stage where genome-wide patterns of DNA methylation are being deposited and cell potency

restriction is being established following implantation of the blastocyst. This indicates a previously uncharacterised role for *Lsh* is contributing to global DNA methylation patterns during the early post-implantation stages of development.

5.3.3 Differential regulation of DNA methylation at satellite repeats is dependent on developmental context

The analysis of the 5-mC status of major and minor satellites revealed that DNA methylation of these repetitive elements is differentially regulated by *Lsh* in different cellular contexts. In the absence of *Lsh* in 2i mESCs, minor satellites were severely hypomethylated, whereas there was no effect on DNA methylation levels at major satellites. Conversely, DNA methylation levels were drastically reduced at major satellites in *lsh*$^{-/-}$ EpiSCs compared to E14, and moderately reduced in EBs. This uncovers an interesting feature of *Lsh* function in differentially influencing DNA methylation at specific genomic loci in different developmental contexts, shedding light on the mechanisms underlying repeat methylation during the early stages of development.

5.3.4 Role for Lsh in maintenance methylation?

The extent of DNA hypomethylation at major and minor satellites in the absence of *Lsh* is too considerable to be explained by defective *de novo* DNA methylation alone. Indeed, the level of 5-mC at major satellites in *lsh*$^{-/-}$ EpiSCs is far below that of the 2i mESCs from which they were differentiated, suggesting that not only has there been a lack of *de novo* methylation at these sequences during differentiation to EpiSCs, DNA methylation has actually been lost from these regions during this transition. Furthermore, the severe hypomethylation observed at minor satellites in the absence of *Lsh* is specific to 2i mESCs, where *de novo* DNA methylation is largely attenuated due to transcriptional downregulation of the *de novo* DNMTs. This leads to the assumption that the DNA hypomethylation observed at these repetitive elements arises due to impaired maintenance methylation and progressive loss of 5-mC over subsequent cell divisions. Therefore, these analyses provide further insights into the potential role for *Lsh* beyond its involvement in *de novo* DNA methylation during development.

5.3.5 Lsh influences expression of developmental genes in EpiSCs

Comparative analysis of E14 and *lsh*$^{-/-}$ EpiSC transcriptional profiles uncovered nearly 2000 genes that are up or down-regulated more than two-fold in the absence of *Lsh*. GO term analysis of the biological processes in which these genes are involved revealed an enrichment in terms associated with cell and organism development for both up- and down-regulated genes. In particular, there was an enrichment in terms associated with the specification, differentiation and development of neurons in the GO analysis of upregulated genes. This demonstrates the de-regulation of pathways involved in neuronal development in the absence of *Lsh*, implicating *Lsh* in the processes of neuron lineage specification and development. Indeed, absence of *Lsh* has been previously reported to affect expression of neuron lineage-associated markers during differentiation, influencing the propensity of *lsh*$^{-/-}$ iPSCs and ESCs to differentiate towards this lineage (Yu, Briones, *et al.*, 2014; Han *et al.*, 2017).

These results indicate that the function of *Lsh* influences expression of developmental genes in this culture model of post-implantation development. This could have implications for the specification and establishment of germ layers in the developmental stages subsequent to the post-implantation epiblast. This analysis highlights an intriguing possible function for *Lsh*, however further investigation is required to deduce the effect that *Lsh* has on lineage-specific factor expression and developmental progression.

5.3.6 Lsh expression is enriched in pluripotent cell states

Expression analysis of *Lsh* across the various pluripotent and differentiated culture models used in this study revealed that *Lsh* is most highly expressed in pluripotent culture states, is transcriptionally downregulated upon transition to EpiSCs, and is further reduced following differentiation to EBs. *Lsh* is most highly expressed in serum E14 mESCs, where it is almost two-fold upregulated compared to 2i mESCs. This appears to be counter-intuitive, as absence of *Lsh* had no apparent effect on global or local satellite DNA methylation levels in serum mESCs. One possible explanation for this is that the low levels of *Lsh* present in differentiated cells are sufficient to influence DNA methylation levels in these cellular contexts, and that the higher expression of

Lsh observed in pluripotent cells is related to an alternative function for *Lsh* in these systems. An alternative explanation is that *Lsh* acts indirectly to influence DNA methylation levels in differentiated EpiSCs and EBs. In this case, the higher *Lsh* expression observed in 2i mESCs, and particularly in serum mESCs which supposedly represent a slightly later developmental state, could act to prepare genomic regions that will subsequently become cytosine methylated upon transition to post-implantation stages of development. This preparation could involve a mechanism whereby chromatin state is altered, potentially through the putative chromatin helicase activity of *Lsh*, or through modification of histone modifications. This altered chromatin state in the epiblast prior to implantation could subsequently influence the genomic regions that acquire 5-mC during the wave of *de novo* DNA methylation during implantation of the blastocyst. This potential function of *Lsh* will be discussed in greater detail in the General discussion (Section 6.6) of the work presented in this thesis.

Chapter 6. General discussion

The overarching purpose of the work undertaken in this thesis was to provide insight into the fundamental, yet not fully understood, epigenetic mechanisms involved in the transition from a pluripotent to a more developmentally primed cellular state during the early stages of murine embryonic development. Specifically, I aimed to evaluate the involvement of selected factors in the establishment of DNA methylation during the cellular transitions that accompany the progression towards lineage commitment during early embryogenesis. I predominantly focussed on the function of DNA methylation co-factor and putative chromatin remodelling helicase *Lsh* in these cellular transitions, utilising *in vitro* culture systems to model the developmental stages encompassing the reduction in cellular potential and the concomitant increase in global DNA methylation establishment. Using these culture models that represented pluripotent and developmentally primed states enabled investigation into the impact of absence of *Lsh* on transitions between cellular states and, importantly, the effect on the genome-wide establishment of DNA methylation that accompanies these transitions.

Firstly, I optimised a culture system in which to study the establishment of *de novo* DNA methylation during the transition of E14 and *lsh*$^{-/-}$ mESCs between two pluripotent culture states. The analysis of expression changes of a few key pluripotency and lineage-associated genes, alongside the quantification of global DNA methylation levels, suggested that both E14 and *lsh*$^{-/-}$ mESCs can effectively transition between the hypomethylated ground state of pluripotency (2i) and the hypermethylated, primed state of pluripotency (serum). This indication was reinforced by analysis of genome-wide transcription and DNA methylation profiles which further demonstrated the ability of both E14 and *lsh*$^{-/-}$ mESCs to interconvert between the pluripotent serum and 2i culture states. Crucially, and somewhat surprisingly, these global and local analyses revealed that the absence of *Lsh* does not affect the re-establishment of DNA methylation patterns during reversion from the pluripotent ground state, suggesting that *Lsh* does not contribute to *de novo* DNA methylation in a pluripotent context.

To further investigate the role of *Lsh* in the establishment of DNA methylation that occurs during the restriction of cell fate potential associated with differentiation and germ layer specification, I used culture models representing later developmental stages. E14 and *lsh*$^{-/-}$ 2i mESCs were differentiated to EpiSCs and EBs, which revealed a critical requirement for *Lsh* in contributing to DNA methylation during transition to these cell types, as DNA methylation levels were significantly reduced globally and at major and minor satellites in *lsh*$^{-/-}$ EpiSCs and EBs. A particularly striking decrease in DNA methylation was evident at the pericentromeric major satellites in *lsh*$^{-/-}$ EpiSCs, highlighting an important function for *Lsh* in contributing to DNA methylation at these repetitive regions in a culture model of post-implantation development. This is consistent with the observation of DNA hypomethylation of satellite repeats first described in mutant *lsh*$^{-/-}$ mouse models and subsequently HELLS-dependent ICF syndrome (Dennis *et al.*, 2001; Sun, David W Lee, *et al.*, 2004; De La Fuente *et al.*, 2006; Dunican *et al.*, 2013; Ren *et al.*, 2015; Thijssen *et al.*, 2015).

These investigations into DNA methylation levels globally and at satellite repeats in the absence of *Lsh* also revealed a reduction in DNA methylation globally and at minor satellites in 2i *lsh*$^{-/-}$ mESCs, where *de novo* DNMT activity is low. This provides evidence for a relatively uncharacterised role for *Lsh* in maintenance of DNA methylation at specific genomic loci, dependent on the cellular context. Overall, the findings presented in this thesis have resulted in the following main conclusions (also summarised in Figure 6.1):

- *Lsh* contributes to DNA methylation globally and at satellite repeats in culture models representing ground state pluripotency and post-implantation development
- *Lsh* differentially regulates DNA methylation at major and minor satellites depending on cellular context

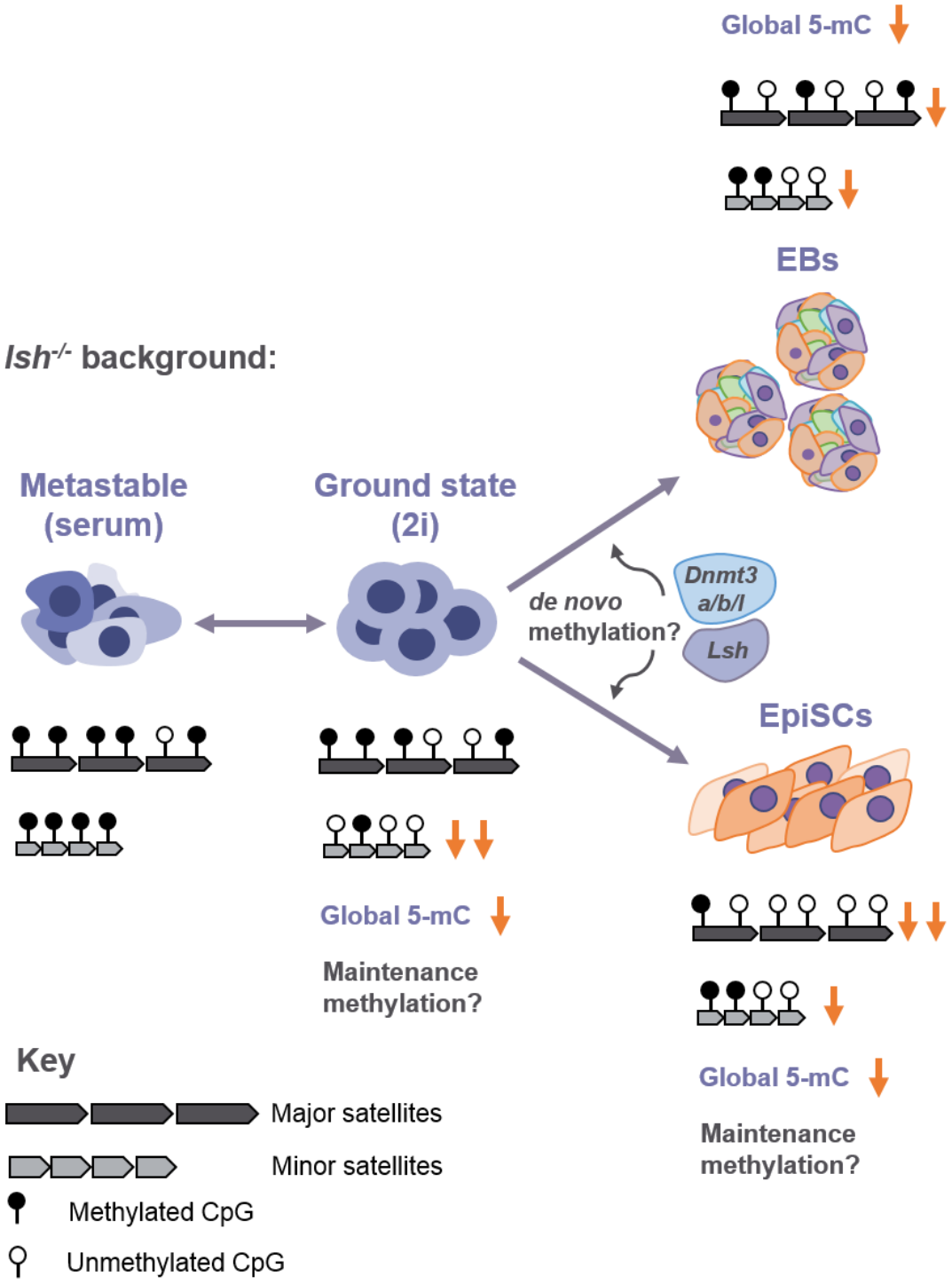


Figure 6.1. Schematic summarising the main findings of this thesis. The methylation status of major and minor satellites in *lsh*$^{-/-}$ serum and 2i mESCs, EpiSCs and EBs is depicted diagrammatically, with the orange arrows indicating DNA hypomethylation at these sequences when compared to the equivalent E14 cells. One arrow indicates moderate DNA hypomethylation, while two arrows indicates severe hypomethylation. A reduction of global DNA methylation in *lsh*$^{-/-}$ cultures compared to E14 is also indicated in each culture state. The possible involvement of *Lsh* in *de novo* or maintenance methylation within culture states or during transition to EBs or EpiSCs is also indicated.

6.1 Technical considerations

The *lsh*$^{-/-}$ mESC line used throughout this study was generated using the first iteration of CRISPR/Cas9 gene editing methodology. This involves targeting WT Cas9 nuclease to the desired genomic locus using a sgRNA, which is engineered to contain an approximately 20-base sequence complementary to the genomic DNA target, next to a PAM site. This induces a double strand break at the intended locus, which is repaired by non-homologous end-joining (NHEJ). This error-prone repair process leads to the introduction of insertion/deletion mutations (indels), which can disrupt the translational reading frame of a coding sequence, resulting in loss of gene function (Ran *et al.*, 2013; reviewed by Sander and Joung, 2014). This technique has provided a valuable tool to quickly and effectively edit the genome of a range of eukaryotic cell culture systems. However, CRISPR/Cas9 gene editing using this method is also associated with the occurrence of off-target mutations, where Cas9 is directed to unintended genomic targets that show high sequence similarity to the desired target (Tsai and Joung, 2016).

Towards the end of the study presented in this thesis, an off-target mutation was discovered in the CRISPR-generated *lsh*$^{-/-}$ mESCs. The top two genic and top two intergenic off-target sites (as predicted by the CRISPR sgRNA design software and outlined in Table 6.1) were inspected for off-target mutations using Sanger sequencing. I found no mutations in the top two intergenic sites and the *Myh7b* gene (data not shown). However, a 17 bp deletion at the site of the *Tbkbp1* gene was uncovered (Figure 6.2 A). This mutation was an out-of-frame deletion in exon 3 of the gene, resulting in a premature stop codon. This gene, also known as *Sintbad*, encodes an adaptor protein which is proposed to be involved in interferon signalling during the innate immune response to viral infection (Ryzhakov and Randow, 2007). I checked the expression of this gene in serum and 2i mESCs and EpiSCs using the RNA-seq datasets generated from these samples. Compared to *Lsh* expression as a reference, *Tbkbp1* is very lowly expressed in the culture models utilised throughout this study (Figure 6.2 B).

Chromosomal region	Gene	No. of mismatches compared to guide
chr11:97,010,011-97,010,033	*Tbkbp1*	2
chr11:107,604,794-107,604,816	N/A	2
chr16:98,203,202-98,203,224	N/A	4
chr2:155,454,888-155,454,910	*Myh7b*	3

Table 6.1. Details of genomic regions predicted to be the most likely unintended targets of the *Lsh* sgRNA used to generate *Lsh*$^{-/-}$ mESCs.

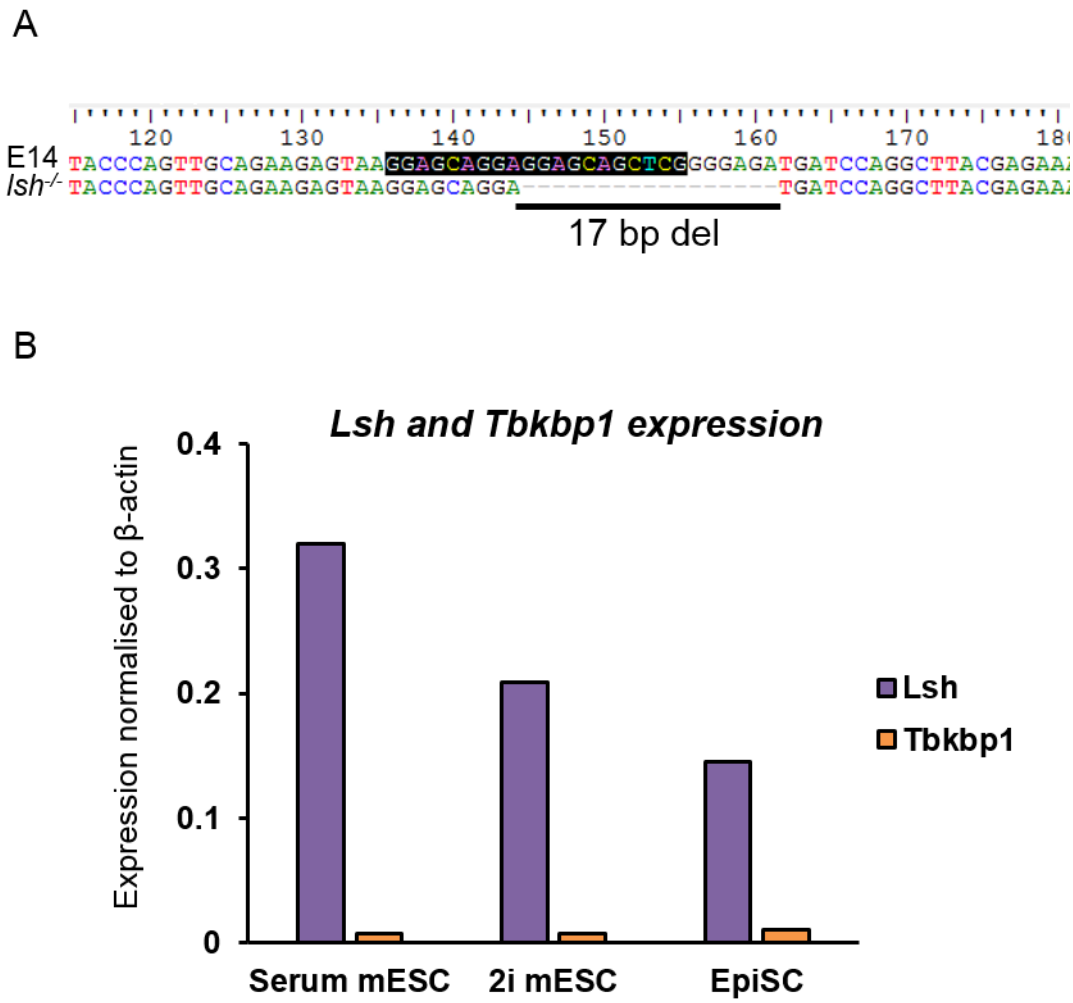


Figure 6.2. Off-target mutation in CRISPR-generated *Lsh*$^{-/-}$ mESCs. A. Sanger sequencing analysis showing alignment of *Tbkbp1* exon 3 in E14 and *Lsh*$^{-/-}$ mESCs, revealing a 17 bp deletion in *Lsh*$^{-/-}$ mESCs. Highlighted region indicates the 20 bp sequence showing similarity to the intended target of the sgRNA in *Lsh* exon 1. B. RNA-seq data showing mRNA expression levels of *Lsh* and *Tbkbp1* in serum and 2i E14 mESCs and EpiSCs.

Since the *Tbkbp1* gene is not thought to be involved in any DNA methylation mechanisms and is lowly expressed in the mESC culture models used in this study, the off-target mutation in this gene should not have impacted any of the results generated in this study. Furthermore, the global and repeat DNA hypomethylation that I have uncovered using the CRISPR-generated *lsh*$^{-/-}$ mESCs has been previously reported in *Lsh*-depleted murine embryonic tissues and somatic cells, validating the findings presented in this thesis. However, further experiments should also be performed in the future to confirm that this off-target mutation did not affect any of the results generated throughout this project. A rescue experiment where *Lsh* is re-expressed in *lsh*$^{-/-}$ mESCs prior to 2i adaptation and differentiation to EpiSCs and EBs would verify that the observed DNA hypomethylation and gene mis-expression is caused by the ablation of *Lsh*. These findings could be further validated using an independently generated *lsh*$^{-/-}$ mESC line, as was intended by the derivation of mESCs from *lsh*$^{-/-}$ mouse blastocysts described in Chapter 3, Section 3.2.2. Overall, although further experiments could corroborate the findings presented in this thesis and fully mitigate the uncertainties associated with the presence of an off-target mutation in the CRISPR-generated *lsh*$^{-/-}$ mESCs, the fact that *Tbkbp1* is very lowly expressed in the culture models I use and is reported to function in an entirely different cellular process to DNA methylation establishment suggests that a mutation in this gene does not appear to contribute to the results generated throughout this study. However, it serves as a warning that CRISPR technology can produce unintentional off-target gene editing that can complicate mutational analysis.

6.2 *Lsh* contribution to DNA methylation during early development

Lsh has been convincingly demonstrated to contribute to DNA methylation during development, as its deletion in mice results in substantial DNA hypomethylation globally and at specific genomic loci in embryonic tissues and somatic cells (Dennis *et al.*, 2001; Sun, David W Lee, *et al.*, 2004; Zhu *et al.*, 2006; Tao *et al.*, 2011; Yu, Briones, *et al.*, 2014). Due to the high expression of *Lsh* in pluripotent cells and its proposed function in *de novo* DNA methylation, I hypothesised that *Lsh* contributes to the wave of *de novo* DNA methylation that occurs in the early stages of development during implantation of the blastocyst (Zhu *et al.*, 2006; Xi *et al.*, 2009; Termanis *et al.*,

2016). Initially, I used the genome-wide re-establishment of DNA methylation that occurs during reversion from the hypomethylated ground state to investigate the role of *Lsh* in *de novo* DNA methylation in a pluripotent context reminiscent of pre-implantation development. It was surprising to discover that the absence of *Lsh* had no impact on the global re-establishment of DNA methylation that occurs during reversion from the pluripotent ground state. Genome-wide analysis of serum and reversion E14 and *lsh*$^{-/-}$ mESC DNA methylation profiles further revealed that lack of *Lsh* did not affect the distribution of DNA methylation following reversion, but rather emphasised the high degree of similarity of 5-mC profiles between reverted and serum mESCs in both E14 and *lsh*$^{-/-}$ backgrounds. These experiments lead to the initial conclusion that, despite its high expression in pluripotent cells, *Lsh* does not contribute to DNA methylation establishment in a pluripotent context. This also implies that *Lsh* does not partake in the early stages of *de novo* DNA methylation establishment in the implanting blastocyst.

The adoption of this culture system to examine the involvement of other epigenetic factors in the *de novo* re-methylation of the genome revealed that *Dnmt3l* is required for DNA methylation establishment during transition from 2i to serum culture. This indicates that *Dnmt3l* is required for *de novo* methylation initiating from the pluripotent ground state, consistent with its essential function in *de novo* DNA methylation and silencing of genomic imprints and retrotransposons (Bourc'his *et al.*, 2001; Hata *et al.*, 2002; Bourc'his and Bestor, 2004; Kaneda *et al.*, 2004). However, although the DNA hypomethylation apparent in *Lsh*-depleted embryos indicates that *Lsh* contributes to DNA methylation during development, it appears not to be required for the *de novo* DNA methylation that initiates from this early developmental stage. This highlights the distinct signatures of epigenetic factors that are required for DNA methylation establishment in different cellular contexts.

To determine whether *Lsh* facilitates DNA methylation establishment during transition to culture models representing later developmental stages, 2i E14 and *lsh*$^{-/-}$ mESCs were differentiated to EBs and EpiSCs. Global DNA methylation analysis by LC-MS revealed a failure to establish WT levels of DNA methylation in the absence of *Lsh* in both EBs and EpiSCs, indicating the involvement of *Lsh* in *de novo* DNA methylation

establishment that occurs during differentiation of 2i mESCs to culture models reflective of post-implantation stages of development. However, this LC-MS analysis also revealed a significant difference in global 5-mC levels between 2i E14 and *lsh*$^{-/-}$ mESCs prior to differentiation. Indeed, comparison of genome-wide DNA methylation profiles of 2i E14 and *lsh*$^{-/-}$ mESCs highlighted the loss of methylation at more than 30,000 CpGs, most of which reside in intronic and intergenic genomic regions, although a proportion are located within LINE-1 subfamilies. This is consistent with many studies which demonstrate DNA methylation loss at LINE-1 elements in *lsh*$^{-/-}$ embryonic tissues and somatic cells (Dennis *et al.*, 2001; Dunican *et al.*, 2013; Yu, McIntosh, *et al.*, 2014; Ren *et al.*, 2015). However, analysis of the 5-mC profiles of 2i *lsh*$^{-/-}$ mESCs has revealed that hypomethylation of LINE-1 repeats observed in somatic tissues and cells may have persisted from the pre-implantation stage of development. Similarly, analysis of the CpG methylation status of minor satellites by bisulfite sequencing and methylation-sensitive restriction digestion coupled with Southern blotting uncovered a striking loss of 5-mC at these repeats in 2i *lsh*$^{-/-}$ mESCs compared to E14. This minor satellite hypomethylation was partially maintained following differentiation to EpiSCs and EBs, providing further evidence that some of the hypomethylation observed in later developmental stages could have originated from the pre-implantation epiblast.

Although interesting, the reduced level of 5-mC observed globally and at specific genomic regions in 2i *lsh*$^{-/-}$ mESCs made it more difficult to delineate whether the global DNA hypomethylation observed in *lsh*$^{-/-}$ EBs and EpiSCs is a result of a defect in *de novo* DNA methylation in the absence of *Lsh*, or is due to the reduced 5-mC level evident in 2i *lsh*$^{-/-}$ mESCs prior to differentiation. However, bisulfite sequencing and methylation-sensitive Southern blotting analysis of CpG methylation at major satellites showed a similar level of 5-mC at these sequences in 2i E14 and *lsh*$^{-/-}$ mESCs, whereas a striking reduction in major satellite DNA methylation was evident in the absence of *Lsh* following transition to EpiSCs. This demonstrates that, at least at some genomic loci, the DNA hypomethylation observed at later developmental stages is not due to a pre-existing lack of DNA methylation at these regions earlier in development. To further deduce to what extent the global DNA hypomethylation observed in *lsh*$^{-/-}$ EBs and EpiSCs is caused by impaired *de novo* methylation in the absence of *Lsh*, a

regulatable expression vector could be used to rescue *Lsh* expression in 2i *lsh*$^{-/-}$ mESCs. This would presumably also rescue the DNA hypomethylation observed in 2i *lsh*$^{-/-}$ mESCs, meaning that global DNA methylation levels would be comparable between 2i E14 and *lsh*$^{-/-}$ mESCs prior to differentiation. *Lsh* expression could then be down-regulated during differentiation to EpiSCs and EBs, meaning that any DNA hypomethylation observed in *lsh*$^{-/-}$ EpiSCs and EBs could be attributed specifically to the function of *Lsh* in DNA methylation establishment during differentiation.

Overall, the results presented throughout this thesis demonstrate a previously uncharacterised role for *Lsh* in modulating DNA methylation in pre-implantation and post-implantation development. These analyses provide the first characterisation of *Lsh* contribution to DNA methylation in the pluripotent ground state and in an EpiSC culture model. More in-depth analysis of DNA methylation profiles using a technique with increased genome coverage compared to ERRBS would provide further insight into the genomic regions at which DNA methylation is regulated by *Lsh* during early development. Furthermore, temporal regulation of *Lsh* expression using a regulatable expression system would allow dissection of the contribution of *Lsh* to DNA methylation in culture models representing distinct developmental stages.

6.3 A role for *Lsh* in maintenance methylation during development?

Lsh is proposed to predominantly contribute to *de novo* DNA methylation rather than maintenance methylation. *Lsh* was shown to be required for DNA methylation establishment on episomal vectors introduced into MEFs, while it was not required for methylation maintenance of already methylated episomes (Zhu *et al.*, 2006). *Lsh* is thought to facilitate *de novo* DNA methylation primarily through its interaction with *Dnmt3b*. This has been demonstrated during differentiation of mESCs to EBs, where depletion of *Lsh* has been reported to reduce the association of *Dnmt3b* with certain genomic targets such as the *Oct4* gene promoter as well as specific repeat elements, resulting in impaired *de novo* methylation at these regions during differentiation (Xi *et al.*, 2009; Ren *et al.*, 2015). The role of *Lsh* in maintenance methylation is less well characterised. *Lsh* has been shown to interact directly with *Dnmt1* and facilitate its association with chromatin (Myant and Stancheva, 2008; Dunican, Pennings and

Meehan, 2015). However, *Lsh* was also shown to be dispensable for the recruitment of *Dnmt1* to replication foci during S phase (Yan *et al.*, 2003).

The analysis of major and minor satellite DNA methylation in mESCs and EpiSCs has provided further evidence for a role for *Lsh* in maintenance methylation during development. There is a dramatic reduction of 5-mC at minor satellites in 2i *lsh*$^{-/-}$ mESCs compared to E14. Furthermore, the percentage CpG methylation at major satellites in *lsh*$^{-/-}$ EpiSCs is less than half of that of 2i *lsh*$^{-/-}$ mESCs, indicating that not only is there a failure to establish DNA methylation at these repeats during transition to EpiSCs, DNA methylation is lost from these regions during the transition. This suggests that *Lsh* is required for maintenance methylation at these repetitive sequences, although this function is dependent on cellular context. Therefore, it would be interesting to assess the association of *Dnmt1* with these satellite repeats in these different cellular contexts in the presence and absence of *Lsh*. This would provide further insights into the mechanism by which *Lsh* promotes maintenance of DNA methylation at these regions during development.

6.4 Influence of *Lsh* on gene expression during development

Expression analysis of a few key lineage-associated and DNMT genes by qRT-PCR demonstrated that the mRNA levels of these genes in serum *lsh*$^{-/-}$ mESCs represent an intermediate expression level between those observed in serum and 2i E14 mESCs. Further analysis and comparison of the transcriptomes of these three mESC cultures discovered a larger (775) subset of genes that exhibit this intermediate expression level in serum *lsh*$^{-/-}$ mESCs between that of E14 serum and 2i mESCs, with GO term enrichment analysis demonstrating that many of these genes are involved in signalling and sensory perception. This indicates that *Lsh* may act to influence expression of genes associated with developmentally-related processes in a pluripotent context. Indeed, depletion of *Lsh* has been reported to result in de-repression of selected *Hox* genes and stem cell-associated factors in MEFs and EBs (Xi *et al.*, 2007, 2009; Tao *et al.*, 2010).

This initial indication that *Lsh* influences the transcriptional regulation of genes involved in developmental processes was reinforced by GO analysis of the genes that

are mis-expressed in *lsh*$^{-/-}$ EpiSCs compared to E14. This revealed an enrichment in genes involved in cellular differentiation and development as well as genes associated with morphogenetic transformations that occur during development. Interestingly, although *Lsh* is generally associated with transcriptional repression due to its contribution to *de novo* methylation of selected gene promoters (Myant and Stancheva, 2008; Ren *et al.*, 2017), a relatively similar number of genes are downregulated as there are upregulated in *lsh*$^{-/-}$ EpiSCs compared to E14 (904 vs 1017, respectively). Of the genes that were upregulated more than two-fold in *lsh*$^{-/-}$ EpiSCs, there was a significant enrichment of genes involved in neuron fate specification and differentiation, suggesting a skew in differentiation towards the neuronal lineage could occur at later developmental stages. This is consistent with a previous report demonstrating an increased propensity of *lsh*$^{-/-}$ iPSCs to differentiate towards the neuronal lineage, which is accompanied by increased expression of some neuronal genes in *lsh*$^{-/-}$ iPSCs (Yu, Briones, *et al.*, 2014). Therefore, it appears that *Lsh* could contribute to the regulation of key genes involved in development and cell fate decisions, influencing the cellular plasticity and differentiation potential of cells during early development. The generation and analysis of DNA methylation profiles of *lsh*$^{-/-}$ EpiSCs would uncover whether the aberrant transcription patterns evident in these cells are a result of gene promoter or enhancer DNA hypomethylation, or whether *Lsh* influences gene expression through an uncharacterised indirect or DNA methylation-independent mechanism.

The primary target of *Lsh* is repetitive elements, and depletion of *Lsh* has been shown to result in transcriptional reactivation primarily of IAP elements, although de-repression of other repetitive sequences such as minor and major satellites is also reported in embryonic tissues and somatic cells (Sun, David W Lee, *et al.*, 2004; Dunican *et al.*, 2013; Yu, McIntosh, *et al.*, 2014). It would therefore be interesting to assess whether there is transcriptional de-repression of repetitive elements during earlier developmental stages through transcriptome analysis of *lsh*$^{-/-}$ mESCs and EpiSCs. Evaluation of major and minor satellite expression in these contexts would uncover whether a substantial loss of DNA methylation at these repeats results in reactivation of gene expression, or whether redundant silencing mechanisms exist, as

has been described for some repeat sequences such as LINE-1 elements in *lsh*$^{-/-}$ MEFs (Dunican *et al.*, 2013).

6.5 *Lsh* expression during development

The analysis of *Lsh* expression in the culture models I utilised throughout this study demonstrated that *Lsh* is most highly expressed in mESCs, and is downregulated upon transition to EpiSCs, and even more so in EBs. This result was unsurprising, as *Lsh* has been previously reported to be highly expressed in pluripotent cells and downregulated during development, leading to a reduced but ubiquitous expression pattern throughout the mouse embryo and somatic tissues (Raabe *et al.*, 2001; Xi *et al.*, 2009). The human homologue of *Lsh*, HELLS, was also identified as one of 40 genes that is consistently highly expressed across 38 human ESC transcriptome datasets, suggesting that the elevated expression of *Lsh* in a pluripotent context is conserved and is in fact characteristic of ESCs (Assou *et al.*, 2007). This may seem counter-intuitive, as the effect of absence of *Lsh* on global DNA methylation appears to be more prominent at later stages of embryonic development, as a 50% decrease in global 5-mC is evident in *lsh*$^{-/-}$ embryos, compared to the approximate 20% reduction in global DNA methylation levels that I show in *lsh*$^{-/-}$ 2i mESCs, EpiSCs and EBs (Tao *et al.*, 2011; Yu, Briones, *et al.*, 2014). This apparent enhanced DNA hypomethylation at later developmental stages could be explained by impaired maintenance methylation in the absence of *Lsh*. In this scenario, high levels of *Lsh* in WT pluripotent cells could act to facilitate *de novo* DNA methylation during exit from pluripotency, before being down-regulated during differentiation. The lower expression levels apparent in differentiated cells could be sufficient to assist maintenance of DNA methylation throughout subsequent cell divisions. Indeed, the majority of genes identified as drivers and facilitators of exit from pluripotency from a number of large-scale screens have been shown to be more highly expressed in the pluripotent ground state, and are then down-regulated upon exit from pluripotency (Kalkan and Smith, 2014). Another possibility is that the high expression of *Lsh* in a pluripotent context is due to a DNA methylation-independent function of *Lsh*, perhaps linked to the high rate of proliferation of pluripotent cells, as the expression of *Lsh* has been shown to be most

prominent in highly proliferative cells such as T and B lymphocytes (Jarvis *et al.*, 1996; Raabe *et al.*, 2001).

6.6 *Lsh* function in DNA methylation and chromatin formation at repetitive elements

The main genomic regions affected by DNA methylation loss in the absence of *Lsh* are satellite DNA and dispersed repeat elements, to the extent that *Lsh* has been referred to as an 'epigenetic guardian' of repetitive elements (Huang *et al.*, 2004). One of the proposed mechanisms of the regulation of DNA methylation at repeats by *Lsh* involve its putative chromatin remodelling activity (Geiman *et al.*, 2001; Meehan, Pennings and Stancheva, 2001). *Lsh* has been proposed to promote heterochromatin formation through its ATP-dependent nucleosome remodelling activity, acting to regulate access of DNMTs to certain genomic loci (Meehan, Pennings and Stancheva, 2001; Ren *et al.*, 2015, 2017). Recent reports have suggested that the nucleosome remodelling activity of *Lsh* is the primary mechanism by which it functions to influence DNA methylation patterns. Ren *et al.* (2015) demonstrated reduced nucleosome occupancy at IAP and LINE-1 elements following differentiation to EBs in the absence of *Lsh*, which coincides with impaired *de novo* DNA methylation of these sequences during differentiation. A further study from Ren *et al.* (2017) demonstrated the reduction in active histone marks and chromatin accessibility at an engineered *Oct4* promoter upon *Lsh* binding in mESCs. Upon differentiation, the *Oct4* promoter acquires repressive H3K9me3 and DNA methylation and is subsequently transcriptionally silenced. This elucidates a potentially crucial mechanism by which *Lsh* alters DNA methylation establishment during differentiation, by modulating chromatin accessibility in order to prime certain loci for acquisition of a silent heterochromatin environment. It would be intriguing to assess whether this mechanism is applicable generally to endogenous loci in mESCs during differentiation, and to deduce whether this function of *Lsh* accounts for the DNA methylation perturbations observed in EpiSCs in the absence of *Lsh*.

Furthermore, it would be interesting to assess the nucleosome occupancy, as well as the histone modification signature of the major and minor satellites that exhibit striking DNA hypomethylation in $lsh^{-/-}$ 2i mESCs and EpiSCs, as this would provide

mechanistic insights into the regulation of these repetitive elements in different developmental contexts, as well as how this acts to influence DNA methylation levels of these repeats. This could have implications for the disease mechanisms underlying ICF syndrome, in which mutations in HELLS have recently been identified as causative of the disease (Thijssen *et al.*, 2015). A hallmark of ICF syndrome is pericentromeric chromosomal instability due to DNA methylation loss and decondensation of these regions, leading to characteristic rearrangements of chromosomes 1, 9 and 16 (Jeanpierre *et al.*, 1993; Hagleitner *et al.*, 2007). *Dnmt3b* is also known to account for 50% of ICF cases, indicating a common mechanism of action at pericentromeric repeats alongside *Lsh* (Hansen *et al.*, 1999; Xu *et al.*, 1999; Ehrlich *et al.*, 2008). Investigation of the regulation of chromatin structure and DNA methylation by *Lsh* and *Dnmt3b* in early development using the culture models described in this thesis could add mechanistic understanding into the origination and progression of the genomic instability observed in ICF syndrome in the context of early murine embryonic development.

6.7 Concluding remarks

In this study, I have uncovered a role for *Lsh* in modulating DNA methylation in culture models reflective of early developmental stages. However, the mechanism by which *Lsh* acts to facilitate establishment or maintenance of DNA methylation in these culture models remains unclear. This would be the primary focus for future work, with emphasis on further characterisation of the role of *Lsh* in maintenance methylation, and investigation into whether the mechanism by which *Lsh* functions involves chromatin remodelling activities. Finally, evaluation of the impact of epigenetic regulation by *Lsh* on gene expression patterns during early development would provide insight into the influence of *Lsh* on developmental progression. This should add understanding into the mechanisms underlying the multiple physiological defects and postnatal lethality apparent in *Lsh*-deficient mouse models, which highlight the crucial role that *Lsh* plays during embryonic development.

References

Abdurashitov, M. A., Chernukhin, V. A., Gonchar, D. A. and Degtyarev, S. (2009) GlaI digestion of mouse γ-satellite DNA: study of primary structure and ACGT sites methylation, *BMC Genomics*, 10(1), 322.

Abranches, E., Bekman, E., Henrique, D., Lee, S. and Su, Q. (2013) Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression, *PLoS ONE*, 8(3), e59928.

Acevedo, N., Wang, X., Dunn, R. L. and Smith, G. D. (2007) Glycogen synthase kinase-3 regulation of chromatin segregation and cytokinesis in mouse preimplantation embryos, *Molecular Reproduction and Development*, 74(2), 178–188.

Akalin, A., Garrett-Bakelman, F. E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T. A., Huang, Y., Biswas, D., Hess, J. L., Allis, C. D., Roeder, R. G., Valk, P. J. M., Löwenberg, B., Delwel, R., Fernandez, H. F., Paietta, E., Tallman, M. S., Schroth, G. P., *et al.* (2012) Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia, *PLoS Genetics*, 8(6).

Almouzni, G. and Probst, A. V. (2011) Heterochromatin maintenance and establishment: Lessons from the mouse pericentromere, *Nucleus*, 2(5), 332–338.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W. S., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., Encheva, V., Kudo, N. R., Koseki, H., Sasaki, H. and Hajkova, P. (2016) De novo DNA methylation drives 5hmC accumulation in mouse zygotes, *Nature Cell Biology*, 18(2), 225–233.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. and Shirakawa, M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism, *Nature*, 455(7214), 818–821.

Assou, S., Le Carrour, T., Tondeur, S., Ström, S., Gabelle, A., Marty, S., Nadal, L., Pantesco, V., Réme, T., Hugnot, J.-P., Gasca, S., Hovatta, O., Hamamah, S., Klein, B. and De Vos, J. (2007) A Meta-Analysis of Human Embryonic Stem Cells Transcriptome Integrated into a Web-Based Expression Atlas, *Stem Cells*, 25(4), 961–973.

Augui, S., Nora, E. P. and Heard, E. (2011) Regulation of X-chromosome inactivation by the X-inactivation centre, *Nature reviews. Genetics*, 12(6), 429–42.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. a, Thomas, J. O., Allshire, R. C. and Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on

histone H3 by the HP1 chromo domain, *Nature*, 410(6824), 120–124.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F. and Bourc'his, D. (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity, *Science*, 354(6314), 909–912.

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A. and Schübeler, D. (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation, *Nature*, 520(7546), 243–247.

Baylin, S. B. and Jones, P. A. (2011) A decade of exploring the cancer epigenome - biological and translational implications, *Nature reviews. Cancer*, 11(10), 726–34.

Becker, P. B. and Hörz, W. (2002) ATP-Dependent Nucleosome Remodeling, *Annual Review of Biochemistry*, 71, 247–273.

Beddington, R. S. and Robertson, E. J. (1989) An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo, *Development*, 105(4), 733–737.

Bell, A. C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene, *Nature*, 405(6785), 482–485.

van den Berg, D. L., Snoek, T., Mullin, N. P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I. and Poot, R. A. (2010) An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells, *Cell stem cell*, 6(4), 369–381.

ten Berge, D., Koole, W., Fuerer, C., Fish, M., Eroglu, E. and Nusse, R. (2008) Wnt Signaling Mediates Self-Organization and Axis Formation in Embryoid Bodies, *Cell Stem Cell*, 3(5), 508–518.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L. and Lander, E. S. (2006) A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells, *Cell*, 125(2), 315–326.

Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases, *Journal of Molecular Biology*, 203(4), 971–983.

Bird, A. (2007) Perceptions of epigenetics, *Nature*, 447(7143), 396–8.

Bird, A. P. (1984) DNA methylation - how important in gene control?, *Nature*, 307(5951), 503–504.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T. and Weber, M. (2010) Targets and dynamics of promoter DNA methylation during early mouse development, *Nat Genet*, 42(12), 1093–1100.

Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S. and Jacobsen, S. E. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells, *Science*, 317(5845), 1760–4.

Bourc'his, D. and Bestor, T. H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L, *Nature*, 431(7004), 96–99.

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. and Bestor, T. H. (2001) Dnmt3L and the establishment of maternal genomic imprints, *Science*, 294(5551), 2536–2539.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A. and Jaenisch, R. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells, *Nature*, 441(7091), 349–53.

Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines, *Nature*, 309(5965), 255–6.

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. and Vallier, L. (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos, *Nature*, 448(7150), 191–195.

Brown, S. W. (1966) Heterochromatin, *Science*, 151(3709).

Brzeski, J. and Jerzmanowski, A. (2003) Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors, *Journal of Biological Chemistry*, 278(2), 823–828.

Burkert U, von Rüden T, W. E. (1991) Early fetal hematopoietic development from in vitro differentiated embryonic stem cells, *New Biol.*, 3(7), 698–708.

Burlingame, R. W., Love, W. E., Wang, B. C., Hamlin, R., Nguyen, H. X. and Moudrianakis, E. N. (1985) Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3 A., *Science*, 228(4699), 546–553.

Burns, K. H. and Boeke, J. D. (2012) Human transposon tectonics, *Cell*, 149(4), 740–752.

Canham, M. A., Sharov, A. A., Ko, M. S. H. and Brickman, J. M. (2010) Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript, *PLoS Biology*, 8(5).

Canzio, D., Chang, E. Y., Shankar, S., Kuchenbecker, K. M., Simon, M. D., Madhani, H. D., Narlikar, G. J. and Al-Sady, B. (2011) Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly, *Molecular Cell*, 41(1), 67–81.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007) Nanog safeguards pluripotency and mediates germline development, *Nature*, 450(7173), 1230–1234.

Chen, H., Guo, R., Zhang, Q., Guo, H., Yang, M., Wu, Z., Gao, S., Liu, L. and Chen, L. (2015) Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells, *Proceedings of the National Academy of Sciences of the United States of America*, 112(44), E5936–E5943.

Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. and Jaenisch, R. (1998) DNA hypomethylation leads to elevated mutation rates, *Nature*, 395(6697), 89–93.

Choi, J., Huebner, A. J., Clement, K., Walsh, R. M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S. Y., Sharif, J., Rose, C. M., Mohammad, A., Odajima, J., Charron, J., Shioda, T., Gnirke, A., Gygi, S., Koseki, H., *et al.* (2017) Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells, *Nature*, 548(7666), 219–223.

Chuang, L. S. H., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G. L. and Li, B. F. L. (1997) Human DNA (cytosine-5) methyltransferase PCNA complex as a target for p21(WAF1), *Science*, 277(1992), 1996–2000.

Clapier, C. R. and Cairns, B. R. (2009) The biology of chromatin remodeling complexes, *Annual review of biochemistry*, 78, 273–304.

Clemson, C. M., McNeil, J. A., Willard, H. F. and Lawrence, J. B. (1996) XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear/chromosome structure, *Journal of Cell Biology*, 132(3), 259–275.

Cordaux, R. and Batzer, M. A. (2009) The impact of retrotransposons on human genome evolution, *Nature reviews. Genetics*, 10(10), 691–703.

Costa, F. F. (2008) Non-coding RNAs, epigenetics and complexity, *Gene*, 410(1), 9–17.

Crichton, J. H., Dunican, D. S., MacLennan, M., Meehan, R. R. and Adams, I. R.

(2014) Defending the genome from the enemy within: Mechanisms of retrotransposon suppression in the mouse germline, *Cellular and Molecular Life Sciences*, 71(9), 1581–1605.

Csankovszki, G., Nagy, A. and Jaenisch, R. (2001) Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation, *Journal of Cell Biology*, 153(4), 773–783.

Czechanski, A., Byers, C., Greenstein, I., Schrode, N., Donahue, L. R., Hadjantonakis, A.K. and Reinholdt, L. G. (2014) Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains, *Nature protocols*, 9(3), 559–74.

Deaton, A. M. and Bird, A. (2011) CpG islands and the regulation of transcription, *Genes and Development*, 25(10), 1010–1022.

Deichmann, U. (2016) Epigenetics: The origins and evolution of a fashionable topic, *Developmental Biology*, 416(1), 249–254.

Dennis, K., Fan, T., Geiman, T., Yan, Q. and Muegge, K. (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation, *Genes and Development*, 2940–2944.

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. and Kemler, R. (1985) The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium, *Journal of embryology and experimental morphology*, 87(2), 27–45.

Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S. and Smithies, O. (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells, *Nature*, 330(6148), 576–578.

Downing, G. J., Battey Jr., J. F. and Battey, J. F. (2004) Technical assessment of the first 20 years of research using mouse embryonic stem cell lines, *Stem cells*, 22(7), 1168–80.

Du, R. (2015) Methyl-CpG-binding domain proteins: readers of the epigenome, *Epigenomics*, 7(6).

Dunican, D. S., Cruickshanks, H. A., Suzuki, M., Semple, C. A., Davey, T., Arceci, R. J., Greally, J., Adams, I. R. and Meehan, R. R. (2013) Lsh regulates LTR retrotransposon repression independently of Dnmt3b function, *Genome Biology*, R146.

Dunican, D. S., Pennings, S. and Meehan, R. R. (2015) Lsh Is Essential for Maintaining Global DNA Methylation Levels in Amphibia and Fish and Interacts

Directly with Dnmt1, *BioMed Research International*, 2015, 1–12.

Ehrlich, M., Sanchez, C., Shao, C., Nishiyama, R., Kehrl, J., Kuick, R., Kubota, T. and Hanash, S. M. (2008) ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation, *Autoimmunity*, 41(4), 253–271.

Eisen, J. A., Sweder, K. S. and Hanawalt, P. C. (1995) Evolution of the SNF2 family of proteins: Subfamilies with distinct sequences and functions, *Nucleic Acids Research*, 23(14), 2715–2723.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H. and Bergman, Y. (2008) De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes, *Nature structural & molecular biology*, 15(11), 1176–83.

Evans, M. J. and Kaufman, M. H. (1981) Establishment in culture of pluripotential cells from mouse embryos, *Nature*, 292(5819), 154–156.

von Eyss, B., Maaskola, J., Memczak, S., Möllmann, K., Schuetz, A., Loddenkemper, C., Tanh, M.D., Otto, A., Muegge, K., Heinemann, U., Rajewsky, N. and Ziebold, U. (2012) The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation, *The EMBO journal*, 31(4), 972–85.

Fan, T., Hagan, J. P., Kozlov, S. V, Stewart, C. L. and Muegge, K. (2005) Lsh controls silencing of the imprinted Cdkn1c gene, *Development*, 132(4), 635–44.

Fan, T., Schmidtmann, A., Xi, S., Briones, V., Zhu, H., Suh, H. C., Gooya, J., Keller, J. R., Xu, H., Roayaei, J., Anver, M., Ruscetti, S. and Muegge, K. (2008) DNA hypomethylation caused by Lsh deletion promotes erythroleukemia development, *Epigenetics*, 3(3), 134–142.

Fan, T., Yan, Q., Huang, J., Austin, S., Cho, E., Ferris, D. and Muegge, K. (2003) Lsh-deficient murine embryonal fibroblasts show reduced proliferation with signs of abnormal mitosis, *Cancer Research*, 63(15), 4677–4683.

Feinberg, A. P., Koldobskiy, M. A. and Göndör, A. (2016) Epigenetic modulators, modifiers and mediators in cancer aetiology and progression, *Nature Reviews Genetics*, 17(5), 284–299.

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J. H. and Längst, G. (2011) Nucleosomes protect DNA from DNA methylation in vivo and in vitro, *Nucleic Acids Research*, 39(16), 6956–6969.

Ferguson-Smith, A. C. (2011) Genomic imprinting: the emergence of an epigenetic

paradigm, *Nature reviews. Genetics*, 12(8), 565–575.

Ferry, L., Fournier, A., Tsusaka, T., Adelmant, G., Shimazu, T., Matano, S., Kirsh, O., Amouroux, R., Dohmae, N., Suzuki, T., Filion, G. J., Deng, W., de Dieuleveult, M., Fritsch, L., Kudithipudi, S., Jeltsch, A., Leonhardt, H., Hajkova, P., Marto, J. A., *et al.* (2017) Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation, *Molecular Cell*, 67(4), 550–565.

Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y. L., Walter, J., Cook, S. J., Andrews, S., Branco, M. R. and Reik, W. (2013) FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency, *Cell Stem Cell*, 13(3), 351–359.

Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E. and Defossez, P.A. (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription, *Molecular and cellular biology*, 26(1), 169–81.

Franke, V., Ganesh, S., Karlic, R., Malik, R., Pasulka, J., Horvat, F., Kuzman, M., Fulka, H., Cernohorska, M., Urbanova, J., Svobodova, E., Ma, J., Suzuki, Y., Aoki, F., Schultz, R. M., Vlahovicek, K. and Svoboda, P. (2017) Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes, *Genome Research*, 27(8), 1384–1394.

Garcia-Perez, J. L., Widmann, T. J. and Adams, I. R. (2016) The impact of transposable elements on mammalian development, *Development*, 143(22), 4101–4114.

Garrett-Bakelman, F. E., Sheridan, C. K., Kacmarczyk, T. J., Ishii, J., Betel, D., Alonso, A., Mason, C. E., Figueroa, M. E. and Melnick, A. M. (2015) Enhanced Reduced Representation Bisulfite Sequencing for Assessment of DNA Methylation at Base Pair Resolution, *Journal of Visualized Experiments*, 3791(96), e52246.

Geiman, T. M., Durum, S. K. and Muegge, K. (1998) Characterization of gene expression, genomic structure, and chromosomal localization of Hells (Lsh), *Genomics*, 54(3), 477–483.

Geiman, T. M., Tessarollo, L., Anver, M. R., Kopp, J. B., Ward, J. M. and Muegge, K. (2001) Lsh, a SNF2 family member, is required for normal murine development, *Biochimica et Biophysica Acta - General Subjects*, 211–220.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M. and Carell, T. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates, *PLoS ONE*, 5(12), e15367.

Goll, M. G. and Bestor, T. H. (2005) Eukaryotic cytosine methyltransferases, *Annu Rev Biochem*, 481–514.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C., Zhang, X., Golic, K. G., Jacobsen, S. E. and Bestor, T. H. (2006) Methylation of tRNA Asp by the DNA Methyltransferase Homolog Dnmt2, *Science*, 311, 395–398.

Gowher, H., Liebert, K., Hermann, A., Xu, G. and Jeltsch, A. (2005) Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L, *Journal of Biological Chemistry*, 280(14), 13341–13348.

Goyal, R., Reinhardt, R. and Jeltsch, A. (2006) Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase, *Nucleic Acids Research*, 34(4), 1182–1188.

Grabole, N., Tischler, J., Hackett, J. A., Kim, S., Tang, F., Leitch, H. G., Magnúsdóttir, E. and Surani, M. A. (2013) Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation, *Nature*, 14(7), 629–63767.

Gruenbaum, Y., Stein, R., Cedar, H. and Razin, A. (1981) Methylation of CpG sequences in eukaryotic DNA, *FEBS Letters*, 124(1), 67–71.

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A. and Meissner, A. (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling, *Nature protocols*, 6(4), 468–481.

Guo, G., Yang, J., Nichols, J., Hall, J. S., Eyres, I., Mansfield, W. and Smith, A. (2009) Klf4 reverts developmentally programmed restriction of ground state pluripotency, *Development*, 136(7), 1063–9.

Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H. D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N. C., Benedetti, R., Altucci, L., Jansen, J. H., Walter, J., Gut, I. G., Marks, H. and Stunnenberg, H. G. (2013) Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells, *Cell Stem Cell*, 13(3), 360–369.

Hackett, J. A. and Azim Surani, M. (2014) Regulatory principles of pluripotency: From the ground state up, *Cell Stem Cell*, 15(4), 416–430.

Hackett, J. A., Reddington, J. P., Nestor, C. E., Dunican, D. S., Branco, M. R., Reichmann, J., Reik, W., Surani, M. A., Adams, I. R. and Meehan, R. R. (2012) Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline, *Development*, 139(19), 3623–3632.

Hackett, J. A. and Surani, M. A. (2013) DNA methylation dynamics during the mammalian life cycle., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1609), 20110328.

Hagleitner, M. M., Lankester, A., Maraschio, P., Hulten, M., Fryns, J. P., Schuetz, C., Gimelli, G., Davies, E. G., Gennery, A., Belohradsky, B. H., de Groot, R., Gerritsen, E. J. A., Mattina, T., Howard, P. J., Fasth, A., Reisli, I., Furthner, D., Slatter, M. A., Cant, A. J., *et al.* (2007) Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome), *Journal of Medical Genetics*, 45(2), 93–99.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J. and Surani, M. A. (2002) Epigenetic reprogramming in mouse primordial germ cells, *Mechanisms of Development*, 15–23.

Han, D. W., Tapia, N., Joo, J. Y., Greber, B., Araúzo-Bravo, M. J., Bernemann, C., Ko, K., Wu, G., Stehling, M., Do, J. T. and Schöler, H. R. (2010) Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages, *Cell*, 143(4), 617–627.

Han, J. S., Szak, S. T. and Boeke, J. D. (2004) Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes, *Nature*, 429(6989), 268–274.

Han, Y., Ren, J., Lee, E., Xu, X., Yu, W. and Muegge, K. (2017) Lsh/HELLS regulates self-renewal/proliferation of neural stem/progenitor cells, *Scientific Reports*, 7(1), 1136.

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M. R. and Gartler, S. M. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome, *Proceedings of the National Academy of Sciences* , 96(25), 14412–14417.

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M. and Tilghman, S. M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus, *Nature*, 405(6785), 486–489.

Hassani, S. N., Totonchi, M., Sharifi-Zarchi, A., Mollamohammadi, S., Pakzad, M., Moradi, S., Samadian, A., Masoudi, N., Mirshahvaladi, S., Farrokhi, A., Greber, B., Araúzo-Bravo, M. J., Sabour, D., Sadeghi, M., Salekdeh, G. H., Gourabi, H., Schöler, H. R. and Baharvand, H. (2014) Inhibition of TGFB Signaling Promotes Ground State Pluripotency, *Stem Cell Reviews and Reports*, 10(1), 16–30.

Hata, K., Okano, M., Lei, H. and Li, E. (2002) Dnmt3L cooperates with the Dnmt3

family of de novo DNA methyltransferases to establish maternal imprints in mice, *Development*, 129(8), 1983–1993.

Hayashi, K., Lopes, S. M. C. de S., Tang, F. and Surani, M. A. (2008) Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States, *Cell Stem Cell*, 3(4), 391–401.

He, X., Yan, B., Liu, S., Jia, J., Lai, W., Xin, X., Tang, C. E., Luo, D., Tan, T., Jiang, Y., Shi, Y., Liu, Y., Xiao, D., Chen, L., Liu, S., Mao, C., Yin, G., Cheng, Y., Fan, J., *et al.* (2016) Chromatin remodeling factor LSH drives cancer progression by suppressing the activity of fumarate hydratase, *Cancer Research*, 76(19), 5743–5755.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dia, Q., Song, C.-X., Zhang, K., He, C. and Xu, G.-L. (2010) Tet mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA, *Science*, 333(September), 1303–1307.

Heitz, E. (1928) Das heterochromatin der moose, *I Jahrb wiss Bot*, 69, 762–818.

Hellman, A. and Chess, A. (2007) Gene body-specific methylation on the active X chromosome, *Science*, 315(5815), 1141–3.

Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins, *Molecular and cellular biology*, 18(11), 6538–47.

Hill, P. W. S., Amouroux, R. and Hajkova, P. (2014) DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: An emerging complex story, *Genomics*, 104(5), 324–333.

Holliday, R. & Pugh, J. E. (1975) DNA modification mechanisms and gene activity during development, *Science*, 187(4173), 226–232.

Holliday, R. and Pugh, J. E. (1975) DNA Modification Mechanisms and Gene Activity during Development Developmental clocks may depend on the enzymic modification of specific bases in repeated DNA sequences, *Science*, 187, 226–232.

Huang, J., Fan, T., Yan, Q., Zhu, H., Fox, S., Issaq, H. J., Best, L., Gangi, L., Munroe, D. and Muegge, K. (2004) Lsh, an epigenetic guardian of repetitive elements, *Nucleic Acids Research*, 32(17), 5019–5028.

Huang, Y., Osorno, R., Tsakiridis, A. and Wilson, V. (2012) In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation, *Cell Reports*, 2(6), 1571–1578.

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., Humphray, S., Cox, T., Langford, C. and Bird, A. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci, *PLoS Biology*, 6(1), 0037–0051.

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R. and Bird, A. P. (2010) Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome, *PLoS Genet*, 6(9).

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, *Science*, 333(6047), 1300–3.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J. and Ramsahoye, B. (2004) Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells, *Molecular and cellular biology*, 24(20), 8862–71.

Jarvis, C. D., Geiman, T., Vila-Storm, M. P., Osipovich, O., Akella, U., Candeias, S., Nathan, I., Durum, S. K. and Muegge, K. (1996) A novel putative helicase produced in early murine lymphocytes, *Gene*, 169(2), 203–207.

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A. and Viegas-pequignot, E. (1993) An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome, *Human Molecular Genetics*, 2(6), 731–735.

Jeddeloh, J. A., Stokes, T. L. and Richards, E. J. (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein, *Nature genetics*, 22(1), 94–97.

Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T. P., Lyko, F., Reuter, G., Ankri, S., Nellen, W., Schaefer, M. and Helm, M. (2017) Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation, *RNA Biology*, 14(9), 1108–1123.

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation, *Nature*, 449(7159), 248–51.

Jones, P. A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond, *Nature reviews. Genetics*, 13(7), 484–92.

Kalkan, T. and Smith, A. (2014) Mapping the route from naive pluripotency to lineage specification., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1657), 20130540-.

Kanai, Y., Ushijima, S., Nakanishi, Y., Sakamoto, M. and Hirohashi, S. (2003) Mutation of the DNA methyltransferase (DNMT) 1 gene in human colorectal cancers, *Cancer Letters*, 192(1), 75–82.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E. and Sasaki, H. (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting, *Nature*, 429(6994), 900–903.

Keyes, W. M., Pecoraro, M., Aranda, V., Vernersson-Lindahl, E., Li, W., Vogel, H., Guo, X., Garcia, E. L., Michurina, T. V., Enikolopov, G., Muthuswamy, S. K. and Mills, A. A. (2011) Δnp63α is an oncogene that targets chromatin remodeler Lsh to drive skin stem cell proliferation and tumorigenesis, *Cell Stem Cell*, 8(2), 164–176.

Kim, S. H., Kang, Y. K., Koo, D. B., Kang, M. J., Moon, S. J., Lee, K. K. and Han, Y. M. (2004) Differential DNA methylation reprogramming of various repetitive sequences in mouse preimplantation embryos, *Biochemical and Biophysical Research Communications*, 324(1), 58–63.

Kipling, D., Wilson, H. E., Mitchell, A. R., Taylor, B. A. and Cooke, H. J. (1994) Mouse centromere mapping using oligonucleotide probes that detect variants of the minor satellite, *Chromosoma*, 103(1), 46–55.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T. and Wutz, A. (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation, *PLoS Biology*, 2(7), e171.

Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A. F., Jones, V., Hor, A., De Alencastro, G., Logan, G. J., Teber, E. T., Tam, O. H., Stutz, M. D., Alexander, I. E., Pickett, H. A. and Tam, P. P. L. (2014) The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak, *Cell Stem Cell*, 14(1), 107–120.

Kornberg, R. D. and Thomas, J. O. (1974) Chromatin structure; oligomers of the histones, *Science*, 184(139), 865–868.

Kouzarides, T. (2007) Chromatin Modifications and Their Function, *Cell*, 128(4), 693–705.

Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, *Science*, 324(5929), 929–930.

De La Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I. and Muegge, K. (2006) Lsh is required for meiotic chromosome synapsis and retrotransposon

silencing in female germ cells, *Nat Cell Biol*, 8(12), 1448–1454.

Laget, S., Joulie, M., Le Masson, F., Sasai, N., Christians, E., Pradhan, S., Roberts, R. J. and Defossez, P. A. (2010) The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA, *PLoS ONE*, 5(8).

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J. and Reik, W. (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse, *Genesis*, 35(2), 88–93.

Larsen, F., Gundersen, G., Lopez, R. and Prydz, H. (1992) CpG islands as gene markers in the human genome, *Genomics*, 13(4), 1095–1107.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Sung, K. W. K., Rigoutsos, I., Loring, J. and Wei, C. L. (2010) Dynamic changes in the human methylome during differentiation, *Genome Research*, 20(3), 320–331.

Lee, D. W., Zhang, K., Ning, Z. Q., Raabe, E. H., Tintner, S., Wieland, R., Wilkins, B. J., Kim, J. M., Blough, R. I. and Arceci, R. J. (2000) Proliferation-associated SNF2-like gene (PASG): A SNF2 family member altered in leukemia, *Cancer Research*, 60(13), 3612–3622.

Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. and Peters, A. H. F. M. (2003) Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin, *Current Biology*, 13(14), 1192–1200.

Lei, H., Oh, S. P., Okano, M., Jüttermann, R., Goss, K. a, Jaenisch, R. and Li, E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells, *Development*, 122(10), 3195–3205.

Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J. G., Smith, A., Surani, M. A. and Hajkova, P. (2013) Naive pluripotency is associated with global DNA hypomethylation, *Nature structural & molecular biology*, 20(3), 311–6.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA, *Cell*, 69(6), 905–14.

Ley, T. J., Ding, L., Walter, M. J., Mclellan, M. D., Lamprecht, T., Larson, D. E., Ph, D., Kandoth, C., Payton, J. E., Baty, J., Welch, J., Harris, C. C., Lichti, C. F., Townsend, R. R., Fulton, R. S., Dooling, D. J., Koboldt, D. C., Schmidt, H., Zhang, Q., *et al.* (2010) Mutations in Acute Myeloid Leukemia, *The New England Journal of*

Medicine, 365(25), 2424–2433.

Li, E., Beard, C. and Jaenisch, R. (1993) Role for DNA methylation in genomic imprinting, *Nature*, 366(6453), 362–5.

Li, E., Bestor, T. H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality, *Cell*, 69(6), 915–926.

Li, M. and Belmonte, J. C. I. (2017) Ground rules of the pluripotency gene regulatory network, *Nature reviews. Genetics*, 18(3), 180–191.

Lin, S. P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J. and Ferguson-Smith, A. C. (2003) Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12, *Nat Genet*, 35(1), 97–102.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B. and Ecker, J. R. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences, *Nature*, 462(7271), 315–22.

Litwin, I., Bakowski, T., Maciaszczyk-Dziubinska, E. and Wysocki, R. (2017) The LSH/HELLS homolog Irc5 contributes to cohesin association with chromatin in yeast, *Nucleic Acids Research*, 45(11), 6404–6416.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H. and Wong, J. (2013) UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9, *Nature communications*, 4, 1563.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution, *Nature*, 389(6648), 251–260.

Lungu, C., Muegge, K., Jeltsch, A. and Jurkowska, R. Z. (2015) An ATPase-deficient variant of the SNF2 family member HELLS shows altered dynamics at pericentromeric heterochromatin, *Journal of Molecular Biology*, 427(10), 1903–1915.

Lynch, M. D., Smith, A. J. H., De Gobbi, M., Flenley, M., Hughes, J. R., Vernimmen, D., Ayyub, H., Sharpe, J. A., Sloane-Stanley, J. A., Sutherland, L., Meek, S., Burdon, T., Gibbons, R. J., Garrick, D. and Higgs, D. R. (2012) An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment, *The EMBO Journal*, 31(2), 317–329.

Mak, W., Nesterova, T. B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P.

and Brockdorff, N. (2004) Reactivation of the Paternal X Chromosome in Early Mouse Embryos, *Science*, 303(5658).

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C. L., Dynlacht, B. D. and Reinberg, D. (2008) Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms, *Molecular Cell*, 32(4), 503–518.

Margueron, R. and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life, *Nature*, 469(7330), 343–9.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A. and Stunnenberg, H. G. (2012) The transcriptional and epigenomic foundations of ground state pluripotency, *Cell*, 590–604.

Marks, H. and Stunnenberg, H. G. (2014) Transcription regulation and chromatin structure in the pluripotent ground state, *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 129–137.

Martens, J. H. A., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P. and Jenuwein, T. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome, *The EMBO journal*, 800–812.

Martin, G. R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells, *Proceedings of the National Academy of Sciences of the United States of America*, 78(12), 7634–7638.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., *et al.* (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters, *Nature*, 466(7303), 253–7.

Meehan, R., Lewis, J. D. and Bird, A. P. (1992) Characterization of MECP2, a vertebrate DNA binding protein with affinity for methylated DNA, *Nucleic Acids Research*, 20(19), 5085–5092.

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. and Bird, A. P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs, *Cell*, 58(3), 499–507.

Meehan, R. R., Pennings, S. and Stancheva, I. (2001) Lashings of DNA methylation, forkfuls of chromatin remodeling, *Genes and Development*, 3231–3236.

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S. and Jaenisch, R.

(2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis, *Nucleic Acids Res.*, 5868–5877.

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S. and Jaenisch, R. (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis, *Nucleic Acids Research*, 5868–5877.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R. and Lander, E. S. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells, *Nature*, 454(7205), 766–770.

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N. Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., Kroeze, L. I., Nesterova, T. B., Jansen, J. H., Xie, H., He, C., Reik, W. and Stunnenberg, H. G. (2016) Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells, *Molecular Cell*, 62(6), 848–861.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, *Nature*, 448(7153), 553–560.

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M. and Schübeler, D. (2008) Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors, *Molecular Cell*, 30(6), 755–766.

Murry, C. E. and Keller, G. (2008) Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development, *Cell*, 132(4), 661–680.

Myant, K. and Stancheva, I. (2008) LSH cooperates with DNA methyltransferases to repress transcription, *Molecular and cellular biology*, 28(1), 215–26.

Myant, K., Termanis, A., Sundaram, A. Y. M., Boe, T., Li, C., Merusi, C., Burrage, J., De Las Heras, J. I. and Stancheva, I. (2011) LSH and G9a/GLP complex are required for developmentally programmed DNA methylation, *Genome Research*, 21(1), 83–94.

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R. and Fraser, P. (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin, *Science*, 322(5908), 1717–20.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex, *Nature*, 393(6683), 386–389.

Nashun, B., Hill, P. W. and Hajkova, P. (2015) Reprogramming of cell fate: epigenetic memory and the erasure of memories past, *The EMBO Journal*, 34(10), 1296–1308.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H. G. and Oliviero, S. (2013) Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs, *Cell*, 155(1).

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F. and Oliviero, S. (2017) Intragenic DNA methylation prevents spurious transcription initiation, *Nature*, 543(7643).

Nestor, C. E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J. M., Harrison, D. J. and Meehan, R. R. (2012) Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes, *Genome Research*, 22(3), 467–477.

Ng, H. H., Jeppesen, P. and Bird, a (2000) Active repression of methylated genes by the chromosomal protein MBD1, *Molecular and cellular biology*, 20(4), 1394–1406.

Ng, R. K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W. and Hemberger, M. (2008) Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5, *Nature cell biology*, 10(11), 1280–1290.

Nichols, J. and Smith, A. (2009) Naive and Primed Pluripotent States, *Cell Stem Cell*, 4(6), 487–492.

Niwa, H., Burdon, T., Chambers, I. and Smith, A. (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3, *Genes & development*, 12(13), 2048–60.

Niwa, H., Ogawa, K., Shimosato, D. and Adachi, K. (2009) A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells, *Nature*, 460(7251), 118–22.

Noma, K., Allis, C. D. and Grewal, S. I. S. (2001) Transitions in Distinct Histone H3 Methylation Patterns at the Heterochromatin Domain Boundaries, *Science*, 293(5532).

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, *Cell*, 247–257.

Okano, M., Xie, S. and Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nature*, 19(3), 219–220.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D., Cheng, X. and Bestor, T. H. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA, *Nature*, 448(7154), 714–717.

Pradhan, S., Bacolla, A., Wells, R. D. and Roberts, R. J. (1999) Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation, *The Journal of biological chemistry*, 274(46), 33002–10.

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A. and Prokhortchouk, E. (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor, *Genes and Development*, 15(13), 1613–1618.

Raabe, E. H., Abdurrahman, L., Behbehani, G. and Arceci, R. J. (2001) An SNF2 factor involved in mammalian development and cellular proliferation, *Developmental Dynamics*, 221(1), 92–105.

Radzisheuskaya, A., Chia, G. L. Bin, dos Santos, R. L., Theunissen, T. W., Castro, L. F. C., Nichols, J. and Silva, J. C. R. (2013) A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages., *Nature cell biology*, 15(6), 579–90.

Raggioli, A., Junghans, D., Rudloff, S. and Kemler, R. (2014) Beta-catenin is vital for the integrity of mouse embryonic stem cells, *PloS one*, 9(1), e86691.

Ragunathan, K., Jih, G. and Moazed, D. (2014) Epigenetic inheritance uncoupled from sequence-specific recruitment, *Science*, 348(6230), 1258699.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P. and Jaenisch, R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a, *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), 5237–5242.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system, *Nat. Protocols*, 2281–2308.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. and Jenuwein, T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases, *Nature*, 406(6796),

593–599.

Reddington, J. P., Pennings, S. and Meehan, R. R. (2013) Non-canonical functions of the DNA methylome in gene regulation., *The Biochemical journal*, 451, 13–23.

Reddington, J. P., Perricone, S. M., Nestor, C. E., Reichmann, J., Youngson, N. A., Suzuki, M., Reinhardt, D., Dunican, D. S., Prendergast, J. G., Mjoseng, H., Ramsahoye, B. H., Whitelaw, E., Greally, J. M., Adams, I. R., Bickmore, W. A. and Meehan, R. R. (2013) Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes, *Genome Biology*, 14(3), R25.

Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M. and Muegge, K. (2015) The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences, *Nucleic acids research*, 43(3), 1444–55.

Ren, J., Hathaway, N. A., Crabtree, G. R. and Muegge, K. (2017) Tethering of Lsh at the Oct4 locus promotes gene repression associated with epigenetic changes, *Epigenetics*, (June), 1-9.

Riggs, A. D. (1975) X inactivation, differentiation, and DNA methylation, *Cytogenetics and cell genetics*, 14(1), 9–25.

Robinson, P. J. and Rhodes, D. (2006) Structure of the “30 nm” chromatin fibre: A key role for the linker histone, *Current Opinion in Structural Biology*, 16(3), 336–343.

Rohde, C., Zhang, Y., Reinhardt, R. and Jeltsch, A. (2010) BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences, *BMC bioinformatics*, 11, 230.

Ryzhakov, G. and Randow, F. (2007) SINTBAD, a novel component of innate antiviral immunity, shares a TBK1-binding domain with NAP1 and TANK, *The EMBO journal*, 26(13), 3180–90.

Sado, T., Fenner, M. H., Tan, S. S., Tam, P., Shioda, T. and Li, E. (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation, *Developmental biology*, 225(2), 294–303.

Sado, T., Okano, M., Li, E. and Sasaki, H. (2004) De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation, *Development*, 131(5).

Saksouk, N., Simboeck, E. and Déjardin, J. (2015) Constitutive heterochromatin formation and transcription in mammals, *Epigenetics & chromatin*, 8, 3.

Sander, J. D. and Joung, J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes, *Nature Biotechnology*, 32(4), 347–355.

Schermelleh, L., Haemmer, A., Spada, F., Rösing, N., Meilinger, D., Rothbauer, U., Cardoso, M. C. and Leonhardt, H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation, *Nucleic Acids Research*, 35(13), 4301–4312.

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. and Rauscher, F. J. (2002) SETDB1: A novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins, *Genes and Development*, 16(8), 919–932.

Senner, C. E., Krueger, F., Oxley, D., Andrews, S. and Hemberger, M. (2012) DNA methylation profiles define stem cell identity and reveal a tight embryonic-extraembryonic lineage boundary, *Stem Cells*, 30(12), 2732–2745.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M. and Koseki, H. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, *Nature*, 450(7171), 908–912.

Sharp, A. J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S. B., Dupre, Y. and Antonarakis, S. E. (2011) DNA methylation profiles of human active and inactive X chromosomes, *Genome Research*, 21(10), 1592–1600.

Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., Brennan, P. M., Baillie, J. K., Collino, A., Ghisletti, S., Sinha, S., Iannelli, F., Radaelli, E., Dos Santos, A., Rapoud, D., Guettier, C., Samuel, D., Natoli, G., Carninci, P., *et al.* (2013) Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma, *Cell*, 153(1), 101–111.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R. and Oberdoerffer, S. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing, *Nature*, 479(7371), 74–9.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides, *Nature*, 336(6200), 688–690.

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A. and Meissner, A. (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo, *Nature*, 339–344.

Sproul, D., Kitchen, R. R., Nestor, C. E., Dixon, J. M., Sims, A. H., Harrison, D. J., Ramsahoye, B. H. and Meehan, R. R. (2012) Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns, *Genome Biology*, 13(10), R84.

Sproul, D. and Meehan, R. R. (2013) Genomic insights into cancer-associated aberrant CpG island hypermethylation, *Briefings in Functional Genomics*, 12(3), 174–190.

Sproul, D., Nestor, C., Culley, J., Dickson, J. H., Dixon, J. M., Harrison, D. J., Meehan, R. R., Sims, A. H. and Ramsahoye, B. H. (2011) Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer, *Proceedings of the National Academy of Sciences*, 108(11), 4364–4369.

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K. and Schübeler, D. (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions, *Nature*, 480(7378), 490–5.

Stein, R., Razin, A. and Cedar, H. (1982) In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells, *Proceedings of the National Academy of Sciences of the United States of America*, 79(11), 3418–22.

Stocking, C. and Kozak, C. A. (2008) Endogenous retroviruses: Murine endogenous retroviruses, *Cellular and Molecular Life Sciences*, 65(21), 3383–3398.

Strahl, B. D. and Allis, C. D. (2000) The language of covalent histone modifications, *Nature*, 403(6765), 41–45.

Sun, L. Q., Lee, D. W., Zhang, Q., Xiao, W., Raabe, E. H., Meeker, A., Miao, D., Huso, D. L. and Arceci, R. J. (2004) Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG, *Genes and Development*, 1035–1046.

Sun, L. Q., Lee, D. W., Zhang, Q., Xiao, W., Raabe, E. H., Meeker, A., Miao, D., Huso, D. L. and Arceci, R. J. (2004) Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG, *Genes and Development*, 18(9), 1035–1046.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T. and Shinkai, Y. (2005) Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9, *Genes and Development*, 19(7), 815–826.

Takahashi, K. and Yamanaka, S. (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors, *Cell*, 126(4), 663–676.

Tao, Y., Xi, S., Briones, V. and Muegge, K. (2010) Lsh mediated RNA Polymerase II stalling at HoxC6 and HoxC8 involves DNA methylation, *PLoS ONE*, 5(2), e9163.

Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., Lee, E. Y., Geiman, T., Huang, J., Stephens, R., Leighty, R. M., Zhao, K. and Muegge, K. (2011) Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences, *Proceedings of the National Academy of Sciences of the United States of America*, 5626–31.

Termanis, A., Torrea, N., Culley, J., Kerr, A., Ramsahoye, B. and Stancheva, I. (2016) The SNF2 family ATPase LSH promotes cell-autonomous de novo DNA methylation in somatic cells, *Nucleic Acids Research*, 44(16), 7592–7604.

Tesar, P. J., Chenoweth, J. G., Brook, F. a, Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells, *Nature*, 448(7150), 196–199.

Thijssen, P. E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., Lemmers, R. J. L. F., de Greef, J. C., Gennery, A., Picco, P., Kloeckener-Gruissem, B., Güngör, T., Reisli, I., Picard, C., Kebaili, K., *et al.* (2015) Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome, *Nature communications*, 6, 7870.

Thoma, F., Koller, T. and Klug, A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent super structures of chromatin, *J.Cell.Biol*, 83(November), 403–427.

Thomas, K. R. and Capecchi, M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells, *Cell*, 51(3), 503–512.

Timp, W. and Feinberg, A. P. (2013) Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host, *Nature Reviews Cancer*, 13(7), 497–510.

Ting, D. T., Lipson, D., Paul, S., Brannigan, B. W., Coffman, E. J., Contino, G., Deshpande, V., John, A., Letovsky, S., Rivera, M. N., Bardeesy, N., Maheswaran, S. and Haber, D. A. (2011) Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers, *Science*, 331(6017), 593–596.

Torres-Padilla, M. E., Chambers, I., Abranches, E., Bekman, E., Henrique, D.,

Ambrosetti, D. C., Basilico, C., Dailey, L., Ashall, L., Horton, C. A., Nelson, D. E., Paszek, P., Harper, C. V., Sillitoe, K., Ryan, S., Spiller, D. G., Unitt, J. F., Broomhead, D. S., Beddington, R. S., *et al.* (2014) Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage, *Development*, 141(11), 2173–81.

Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. and Niwa, H. (2008) Identification and characterization of subpopulations in undifferentiated ES cell culture, *Development*, 909–918.

Trojer, P. and Reinberg, D. (2007) Facultative Heterochromatin: Is There a Distinctive Molecular Signature?, *Molecular Cell*, 28(1), 1–13.

Tsai, S. Q. and Joung, J. K. (2016) Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases, *Nature Reviews Genetics*, 17(5), 300–312.

Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R., Economou, C., Karagianni, E., Zhao, S., Lowell, S. and Wilson, V. (2014) Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors., *Development*, 141(6), 1209–21.

Unoki, M., Nishidate, T. and Nakamura, Y. (2004) ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain, *Oncogene*, 23(46), 7601–10.

Vardimon, L., Kressmann, A., Cedar, H., Maechler, M. and Doerfler, W. (1982) Expression of a cloned adenovirus gene is inhibited by in vitro methylation, *Proceedings of the National Academy of Sciences of the United States of America*, 79(4), 1073–1077.

Veillard, A.-C., Marks, H., Bernardo, A. S., Jouneau, L., Laloë, D., Boulanger, L., Kaan, A., Brochard, V., Tosolini, M., Pedersen, R., Stunnenberg, H. and Jouneau, A. (2014) Stable Methylation at Promoters Distinguishes Epiblast Stem Cells from Embryonic Stem Cells and the In Vivo Epiblasts, *Stem cells and development*, 23(17), 2014–29.

Voigt, P., Tee, W. W. and Reinberg, D. (2013) A double take on bivalent promoters, *Genes and Development*, 27(12), 1318–1338.

Waddington, C. H. (1942) The epigenotype, *Endeavour*, 1, 18–20.

Wallrath, L. L. and Elgin, S. C. R. (1995) Position effect variegation in *Drosophila* is associated with an altered chromatin structure, *Genes and Development*, 9(Heitz 1928), 1263–1277.

Walsh, C., Chaillet, J. and Bestor, T. (1998) Transcription of IAP endogenous

retroviruses is constrained by cytosine methylation, *Nature genetics*, 20(October), 116–117.

Walsh, C. P. and Bestor, T. H. (1999) Cytosine methylation and mammalian development, *Genes & development*, 13(1), 26–34.

Walter, M., Teissandier, A., Pérez-Palacios, R. and Bourc'His, D. (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells, *eLife*, 5(January 2016).

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X. and Liu, J. (2014) Programming and inheritance of parental DNA methylomes in mammals, *Cell*, 157(4), 979–991.

Wang, R., Shi, Y., Chen, L., Jiang, Y., Mao, C., Yan, B., Liu, S., Shan, B., Tao, Y. and Wang, X. (2015) The ratio of FoxA1 to FoxA2 in lung adenocarcinoma is regulated by LncRNA HOTAIR and chromatin remodeling factor LSH, *Scientific Reports*, 5(August), 17826.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome, *Nature*, 420(6915), 520–562.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome, *Nature genetics*, 39(4), 457–66.

Weber, M. and Schübeler, D. (2007) Genomic patterns of DNA methylation: targets and function of an epigenetic mark, *Current Opinion in Cell Biology*, 19(3), 273–280.

Weemaes, C. M., Van Tol, M. J., Wang, J., Van Ostaijen-Ten Dam, M. M., Cja Van Eggermond, M., Thijssen, P. E., Aytekin, C., Brunetti-Pierri, N., Van Der Burg, M., Davies, G., Ferster, A., Furthner, D., Gimelli, G. and Gennery, A. (2013) Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects, *European Journal of Human Genetics*, 21(10), 1219–1225.

Wichterle, H., Lieberam, I., Porter, J. and Jessell, T. M. (2002) Directed differentiation of embryonic stem cells into motor neurons, *Cell*, 110(3), 385–97.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. a, Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988) Myeloid leukaemia

inhibitory factor maintains the developmental potential of embryonic stem cells, *Nature*, 336(6200), 684–687.

Williamson, C. M., Turner, M. D., Ball, S. T., Nottingham, W. T., Glenister, P., Fray, M., Tymowska-Lalanne, Z., Plagge, A., Powles-Glover, N., Kelsey, G., Maconochie, M. and Peters, J. (2006) Identification of an imprinting control region affecting the expression of all transcripts in the Gnas cluster, *Nature Genetics*, 38(3), 350–355.

Winston, F. and Carlson, M. (1992) Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection, *Trends in genetics*, 8(11), 387–391.

Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T. and Migeon, B. R. (1984) Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation, *Proceedings of the National Academy of Sciences of the United States of America*, 81(9), 2806–10.

Wolffe, A. P. and Matzke, M. A. (1999) Epigenetics: regulation through repression, *Science*, 286(5439), 481–486.

Xi, S., Geiman, T. M., Briones, V., Tao, Y. G., Xu, H. and Muegge, K. (2009) Lsh participates in DNA methylation and silencing of stem cell genes, *Stem Cells*, 27(11), 2691–2702.

Xi, S., Zhu, H., Xu, H., Schmidtmann, A., Geiman, T. M. and Muegge, K. (2007) Lsh controls Hox gene silencing during development, *Proc Natl Acad Sci U S A*, 104(36), 14366–14371.

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K. and Li, E. (1999) Cloning, expression and chromosome locations of the human DNMT3 gene family, *Gene*, 236(1), 87–95.

Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J. and Viegas-Péquignot, E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene, *Nature*, 402(6758), 187–191.

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K. and Saitou, M. (2013) PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells, *Cell Stem Cell*, 368–382.

Yan, Q., Cho, E., Lockett, S. and Muegge, K. (2003) Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin, *Molecular and cellular biology*, 23(23), 8416–28.

Yan, X. J., Xu, J., Gu, Z. H., Pan, C. M., Lu, G., Shen, Y., Shi, J. Y., Zhu, Y. M., Tang, L., Zhang, X. W., Liang, W. X., Mi, J. Q., Song, H. D., Li, K. Q., Chen, Z. and Chen, S. J. (2011) Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia, *Nat Genet*, 43(4), 309–315.

Ying, Q. L., Nichols, J., Chambers, I. and Smith, A. (2003) BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3, *Cell*, 115, 281–292.

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, Doble, B., Woodgett, J., P., Cohen, P., and Smith, A. (2008) The ground state of embryonic stem cell self-renewal, *Nature*, 453(May), 519–23.

Yoder, J. A., Soman, N. S., Verdine, G. L. and Bestor, T. H. (1997) DNA (cytosine-5)-methyltransferases in mouse cells and tissues. studies with a mechanism-based probe, *Journal of Molecular Biology*, 270(3), 385–395.

Yoder, J. A., Walsh, C. P. and Bestor, T. H. (1997) Cytosine methylation and the ecology of intragenomic parasites, *Trends in Genetics*, 335–340.

Yu, W., Briones, V., Lister, R., McIntosh, C., Han, Y., Lee, E. Y., Ren, J., Terashima, M., Leighty, R. M., Ecker, J. R. and Muegge, K. (2014) CG hypomethylation in Lsh-/- mouse embryonic fibroblasts is associated with de novo H3K4me1 formation and altered cellular plasticity, *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 5890–5.

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M., Leighty, R., Ecker, J. R. and Muegge, K. (2014) Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways, *Genome research*, 24(10), 1613–23.

Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S. L. and Zilberman, D. (2013) The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin, *Cell*, 153(1), 193–205.

Zeng, W., Baumann, C., Schmidtmann, A., Honaramooz, A., Tang, L., Bondareva, A., Dores, C., Fan, T., Xi, S., Geiman, T., Rathi, R., de Rooij, D., De La Fuente, R., Muegge, K. and Dobrinski, I. (2011) Lymphoid-specific helicase (HELLS) is essential for meiotic progression in mouse spermatocytes, *Biology of reproduction*, 84(6), 1235–41.

Zhang, K., Li, L., Huang, C., Shen, C., Tan, F., Xia, C., Liu, P., Rossant, J. and Jing,

N. (2010) Distinct functions of BMP4 during different stages of mouse ES cell neural commitment, *Development*, 137(13), 2095–105.

Zhang, K., Mosch, K., Fischle, W. and Grewal, S. I. S. (2008) Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin, *Nature structural & molecular biology*, 15(4), 381–8.

Zhao, Q., Zhang, J., Chen, R., Wang, L., Li, B., Cheng, H., Duan, X., Zhu, H., Wei, W., Li, J., Wu, Q., Han, J.-D. J., Yu, W., Gao, S., Li, G. and Wong, J. (2016) Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals, *Nature communications*, 7, 12464.

Zhu, H., Geiman, T. M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E. and Muegge, K. (2006) Lsh is involved in de novo methylation of DNA, *The EMBO journal*, 335–345.

Ziller, M. J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C. B., Bernstein, B. E., Lengauer, T., Gnirke, A. and Meissner, A. (2011) Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types, *PLoS Genetics*, 7(12), e1002389.

Appendix

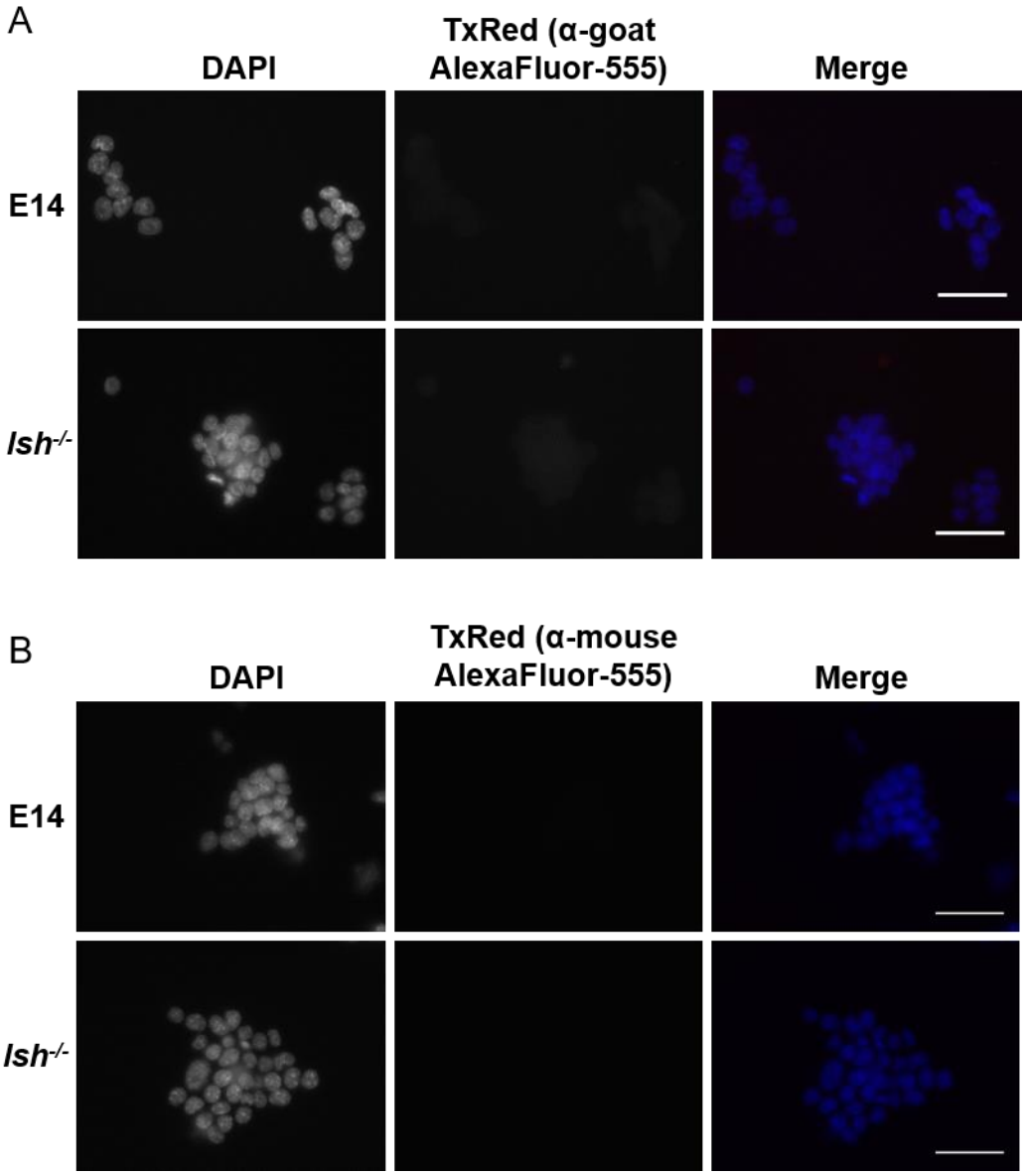
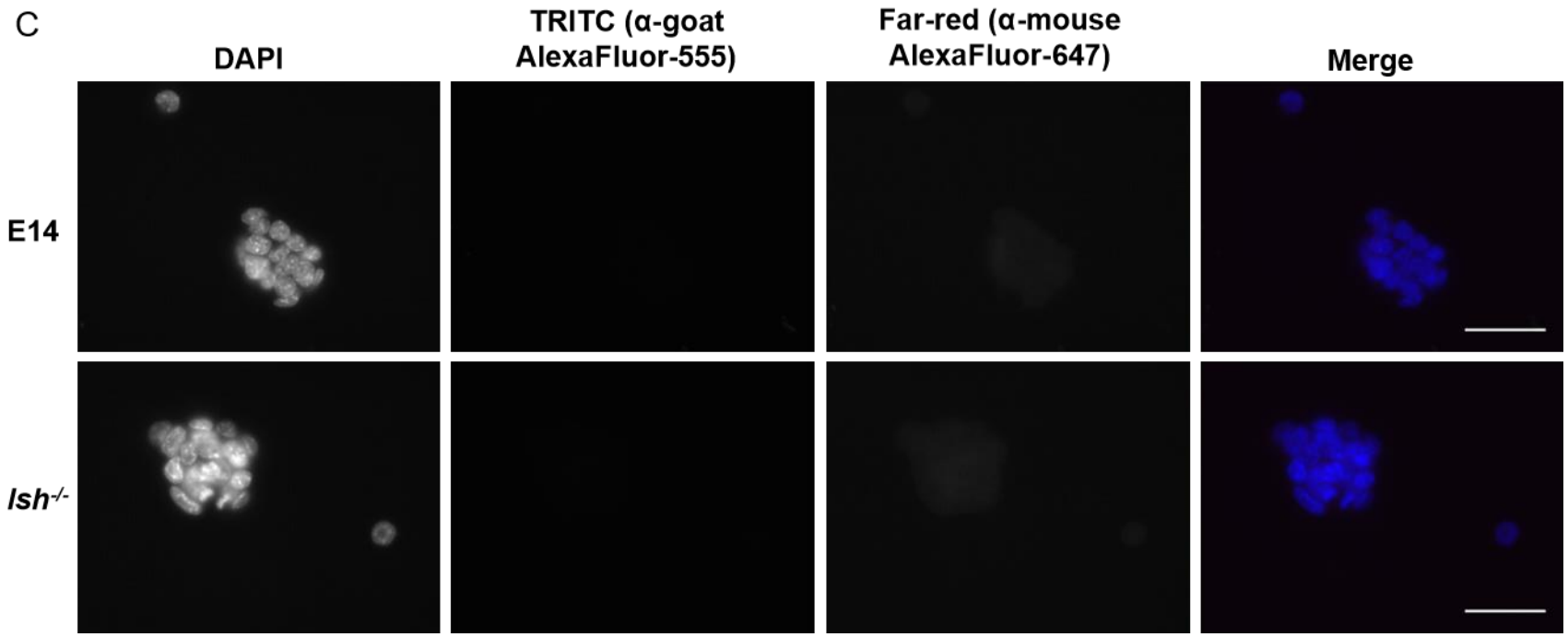


Figure A1. Secondary antibody-only staining. Immunofluorescence on serum E14 and *lsh*$^{-/-}$ mESCs using secondary antibodies only. Each secondary antibody used for immunofluorescence analysis was assessed by secondary antibody-only staining to determine the level of non-specific binding and background fluorescence. The secondary antibody used for *Nanog* staining is shown in (A), the antibody for *Esrrb* staining is shown in (B), and the antibodies used for *Nanog* and *Oct4* co-staining are shown in (C). Scale bars represent 50 µm.


C
DAPI
TRITC (α-goat AlexaFluor-555)
Far-red (α-mouse AlexaFluor-647)
Merge
E14
lsh-/-

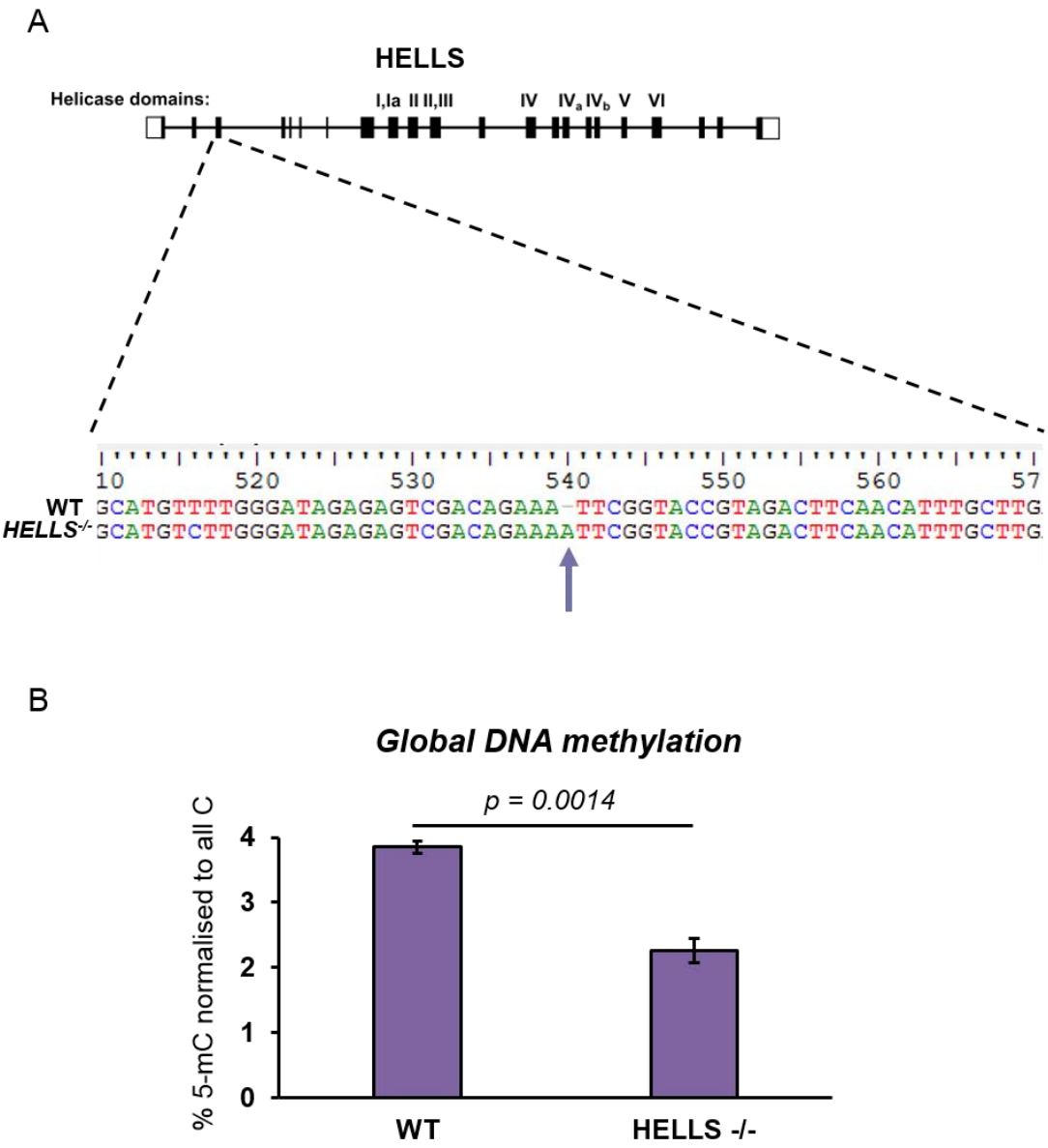


Figure A2. A HELLS mutant human somatic cancer cell line displays substantial global hypomethylation. Analysis of global 5-mC levels in a HELLS mutant human somatic haploid cancer-derived cell line (HAP1) (A) confirmation of HELLS mutation by Sanger sequencing, showing insertion of a single A (indicated by arrow) in exon 3 which creates a premature stop codon. (B) LC-MS analysis of global DNA methylation in WT and HELLS mutant HAP1 cells. Error bars represent +/- standard error of three biological replicates.