

Quantitative analysis of dynamic DNA modifications in pluripotent stem cells and primordial germ cells

Sarah Linnett
Institute of Clinical Sciences
Imperial College London

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration of originality

The research presented in this thesis is my own, unless otherwise stated. All collaborators are duly acknowledged.

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

Many people have supported me throughout my PhD, and I would like to take this opportunity to thank them.

Firstly, thank you Petra for giving me the opportunity to challenge myself. You have taught me the value of perseverance and discipline, and for this I am really grateful.

I would also like to thank the past and present members of the Reprogramming and Chromatin Group. I am incredibly lucky to have had the opportunity to work with so many talented individuals. I will look back on our time in the lab together very fondly. A special thank you is reserved for Rachel and Kirsten. You have guided me through my PhD, and I would have been truly lost without you both. I would also like to thank Tien Chi. I feel extremely fortunate to have completed this journey alongside you. To Harry and Gosia, you have always been so generous with your time and advice. Thank you for keeping me company in tissue culture. Cristina, your help over the past few months has been greatly appreciated. Thank you to Buhe for being so patient with me. I look forwards to seeing your future success as a Group Leader. Peter, you have set an extremely high bar for the rest of us! Thank you for being such a great role model. I would also like to wish Theresa and Karl the best of luck for the remainder of their PhDs. I am sure you will achieve many great things! Finally, thank you to Zelpha and Catherine, the lab would not be able to run smoothly without you.

Additionally, I would like to extend my thanks to the Proteomics Facility. Thank you Peter and Alex, for always being on hand with technical advice and a tool kit.

I would not have made it this far if it wasn't for the unwavering support of my family. Thank you for always being at the end of the phone when I need some reassurance. I regret that I have not been able to see you as much as I would have liked. Hopefully this will change.

My final thanks of course goes to Phil. You have supported me through some really challenging times. I am looking forwards to entering calmer waters and beginning a new adventure together.

Abstract

Almost 65 years have passed since the discovery of 5-methylcytosine (5mC). DNA methylation has since been shown to be involved in gene expression, genomic imprinting, X chromosome inactivation, the repression of retrotransposons and cell differentiation. In the context of mammalian embryonic development, DNA methylation undergoes genome-wide reprogramming in primordial germ cells (PGCs). Despite a huge research effort, the mechanism underlying DNA demethylation in PGCs is unclear. Limited progress is largely due to technical challenges. Many widely used analytical techniques require large amounts of biological material and are therefore inappropriate for studies using small numbers of cells isolated from developing embryos. Our understanding of the mechanisms behind the developmental reprogramming processes may therefore be greatly enhanced by the use of complementary model systems and alternative technologies.

The experiments described herein focus on the development and optimisation of a number of *in vitro* and *in vivo* systems to explore the nature of epigenetic DNA modifications, including 5-methylcytosine, and the mechanisms that dictate dynamic changes in the epigenetic landscape. The development and validation of sensitive LC-MS methods facilitates the quantification of DNA modifications, potentially enables the discovery of novel modifications, and by combining LC-MS with isotope tracing, may fundamentally challenge our current understanding of the stable nature of DNA methylation. Ultimately, we envisage that these approaches will form a versatile and informative platform to study genome-wide epigenetic reprogramming processes in an *in vivo* context.

In many respects, loss of 5-methylcytosine in migratory PGCs mirrors the loss of 5-methylcytosine in mESCs upon transition from serum-containing to 2i medium. Accordingly, I used this accessible system to probe the contribution of DNA methyltransferases, putative DNA demethylases, their cofactors and binding partners to the loss of 5-methylcytosine in nascent PGCs. I show that loss of Dnmt3a isoform 2 is necessary for 2i-induced DNA demethylation. Furthermore, this loss is dependent on RNA-induced silencing. More generally, miRNAs

appear to be necessary for the epigenetic reprogramming of mESCs to the naive ground state. It is possible that the same miRNAs are involved in DNA demethylation in migratory PGCs. I also show that retinol has an unexpected effect on the rate and extent of DNA demethylation. How this finding translates in the context of primordial germ cell development is unknown. It does, however, highlight both a potential source of variation in the published data and the need to understand the interplay between cell metabolism and epigenetic reprogramming.

Table of contents

Declaration of originality	2
Copyright declaration	3
Acknowledgements	4
Abstract	5
Table of contents	7
Frequently used abbreviations	13
List of figures	17
List of tables	19

Chapter 1 An introduction to DNA methylation and the epigenetic regulation of pluripotency and embryonic development

1.1 Understanding biological complexity in the post-genomic era.....	21
1.1.1 The hopes of whole genome sequencing.....	21
1.1.2 Epigenetic regulation increases the genome's information content	23
1.1.3 Gene expression is constrained by the chromatin landscape	23
1.2 DNA methylation and the regulation of cell fate	24
1.2.1 Defining DNA methylation	24
1.2.2 The discovery of the fifth base	24
1.2.3 DNA methylation states are stably inherited	24
1.2.4 The link between DNA methylation and transcriptional repression	25
1.2.5 DNA methylation and lineage restriction in post-implantation development	25
1.3 Structure of the DNA methylome	26
1.3.1 Phylogenetic variation in DNA methylation	26
1.3.2 CpG island promoters	27
1.3.3 The paradox of gene body methylation	29
1.3.4 Non-CpG methylation	30
1.4 DNA methyltransferases	31
1.4.1 Maintenance methylation	31
1.4.2 <i>De novo</i> methylation	32
1.5 The tumultuous search for a putative demethylase	34
1.5.1 Dynamic and static DNA methylation states	34
1.5.2 Passive DNA demethylation	35
1.5.3 Active DNA demethylation in plants	35
1.5.4 The early search for a mammalian demethylase	36
1.5.5 DNA demethylation by direct excision	36
1.5.6 DNA demethylation by deamination and DNA repair	38

1.5.7 DNA demethylation by oxidation and DNA repair	40
1.5.8 The supporting cast	41
1.6 Dynamic DNA methylation during development	45
1.6.1 DNA demethylation during preimplantation development	45
1.6.2 DNA demethylation in pluripotent stem cells	48
1.6.3 DNA demethylation in differentiated cells	49
1.7 Pluripotency	50
1.7.1 Defining the naive state	50
1.7.2 From cleavage to gastrulation	50
1.7.2 Capturing pluripotency - Teratocarcinomas and EC cells	53
1.7.3 Capturing pluripotency - Embryonic stem cells	54
1.7.4 LIF/Stat3/PI3K signalling and pluripotency in embryonic stem cells	55
1.7.5 Akt signalling and pluripotency in embryonic stem cells	57
1.7.6 ERK signalling and defined culture medium	58
1.7.8 Defining the naive ground state	61
1.7.9 Transcriptional control of pluripotency	63
1.7.9.1 Core pluripotency factors	63
1.7.9.2 Ancillary pluripotency factors	64
1.7.10 Prdm14 and the pluripotency network	64

Chapter 2 Materials and Methods

2.1 Mice	68
2.1.1 Mouse husbandry	68
2.1.2 Genotyping	68
2.2 mESC culture protocols	70
2.2.1 Culture conditions	70
2.2.2 mESC lines	71
2.2.3 Passaging	71
2.2.4 Cryopreservation and thawing	72
2.2.5 Cell counting	72
2.2.6 mESC derivation	72
2.2.7 ESC transfection	73
2.2.7.1 Nucleofection	73
2.2.7.2 Puromycin selection	74
2.2.7.3 Optimising Puromycin concentration using a kill curve	74
2.2.7.4 Tamoxifen treatment	74
2.2.8 Preparation of feeder cells	75
2.2.8.1 MEF derivation	75
2.2.8.2 MEF inactivation	75
2.3 Differentiation protocol	76
2.3.1 Myoblast differentiation	76
2.3.1.1 C2C12 medium	76

2.3.1.2 C2C12 differentiation	76
2.3.1.3 Dialysis of horse serum for stable isotope labelling	76
2.3.2 Adipogenesis	77
2.3.2.1 Composition of 3T3-L1 expansion medium abd adipogenic cocktail	77
2.3.2.2. 3T3-L1 adipogenic differentiation protocol	77
2.3.2.3 Oil Red O staining	78
2.3.3 Bone marrow derived macrophage protocols	78
2.3.3.1 Culture medium	78
2.3.3.2 BM Φ derivation	78
2.3.3.3 Dissociation of BM Φ	79
2.3.3.4 Flow cytometric analysis of BM Φ surface markers	79
2.4 Dissection techniques	80
2.4.1 Cortical neuron harvest	80
2.4.2 Nuclei isolation	81
2.4.3 Immunolabelling of nuclei	82
2.4.4 Gonad dissection	82
2.5 Histology	83
2.5.1 Immunostaining of cryosections	83
2.5.2 Immunostaining nucleic acids in cryosections	84
2.5.3 Haematoxylin and Eosin staining	85
2.6 Immunocytology	85
2.6.1 Preparation of glass coverslips	85
2.6.2 Immunofluorescent labelling of cultured cells	86
2.6.3 EdU labelling of adherent mESCs for colocalisation studies	86
2.7 Flow cytometry	87
2.7.1 EdU labelling for flow cytometric studies	87
2.7.2 Flow cytometry of nuclear proteins	87
2.8 Mass spectrometry	88
2.8.1 DNA extraction	88
2.8.2 Purification of digestion enzymes	89
2.8.3 DNA digestion and sample preparation	89
2.8.4 Standard curve preparation	89
2.8.5 Nucleoside quantification	90
2.9 Transcript analysis	91
2.9.1 Isolation of RNA	91
2.9.2 cDNA synthesis	91
2.9.3 Real-time quantitative qPCR	91
2.9.4 Extraction of miRNA from mESCs	93
2.9.5 cDNA synthesis from miRNA enriched RNA	94
2.9.6 miRNA RT-qPCR	94
2.10 Protein expression	95
2.10.1 Protein extraction	95
2.10.2 Protein quantification	95

2.10.3 SDS PAGE and Western Blot	95
2.11 Bacterial protocols	97
2.11.1 Transformation of bacterial cells	97
2.11.2 Bacterial culture and isolation of plasmid DNA	98

Chapter 3 Quantification of global DNA modifications by LC-MS

3.1 Introduction	100
3.1.1 Reverse-phase chromatography	100
3.1.2 Electrospray ionisation	101
3.1.3 Triple quadrupole mass spectrometry	103
3.1.4 Ion suppression, sensitivity and reliability	104
3.2 Experimental aims	107
3.3 Results	108
3.3.1 Prologue	108
I. Selection of a HPLC column based on nucleotide resolution and sensitivity	109
II. Optimising solvent conditions	111
III. Optimisation of sample digestion	112
3.3.2 Tackling ion suppression and enhancement effects	112
3.3.3 Compatibility of common cytological fixatives with nucleoside LC-MS	119
3.4 Discussion	121

Chapter 4 In vivo and ex vivo DNA methylation turnover kinetics

4.1 Introduction	124
4.1.1 Dynamic DNA methylation	124
4.1.2 De novo pathway of pyrimidine biosynthesis	125
4.1.3 Salvage pathway of pyrimidine biosynthesis	127
4.2 Experimental aims	128
4.3 Results	129
4.3.1 Experimental system	129
4.3.2. Detection of isotopomers	129
4.3.3. Stable isotope labelling and quantification of 5mdC turnover in postmitotic cells	133
4.3.4. Optimisation of $^{15}\text{N}_2$ -uridine concentration in proliferating mESCs.....	137
4.3.5. Quantification of isotopomer incorporation in <i>Tet</i> double knockout mESCs	139
4.4. Discussion	144

Chapter 5 Regulation of global DNA methylation in mESCs upon transition to 2i media

5.1 Introduction	148
5.2 Experimental aims	150
5.3 Results	151
5.3.1. Experimental system	151

5.3.2. Dynamics of DNA modifications upon transition of mESCs to 2i culture conditions	152
5.3.3. Transcriptional downregulation of <i>de novo</i> Dnmts drives the rapid onset of DNA de methylation	156
5.3.4. Active and passive mechanisms contribute to DNA demethylation	158
5.3.5. Dynamics of Non-CpG and CpG DNA demethylation	162
5.3.6. Prdm14 transcriptional upregulation is not characteristic of early 2i-mediated epigenetic reprogramming	162
5.3.6. PRDM14 is not solely responsible for Dnmt downregulation	165
5.4 Discussion	166

Chapter 6 miRNAs as regulators of naive and ground state pluripotency

6.1 Introduction	170
6.1.1 MicroRNA	170
6.1.2. miRNA biogenesis	170
6.1.3. miRNA and pluripotency	171
6.3.1.1. The miR-290 cluster	172
6.3.1.2. The miR 302 cluster	173
6.1.4. miRNAs enhance the efficiency of reprogramming	173
6.2 Experimental aims	173
6.3 Results	175
6.3.1. miRNA and the post-transcriptional regulation of <i>de novo</i> Dnmts in ground state ESCs...	175
6.3.2. miRNA regulation of Dnmt3a isoform 2 transcriptional activators and repressors	179
6.3.3. RNA-Seq analysis of <i>Dicer KO mESCs</i>	186
6.3.4. Initial exploratory analysis	187
6.3.5. Genes associated with developmental processes are downregulated in 2i WT and <i>Dicer KO</i> cells	190
6.3.6. <i>Dicer KO</i> cells show reduced expression of naive pluripotency markers	193
6.3.7. <i>Dicer KO</i> cells show characteristics of the epithelial to mesenchymal transition	196
6.3.9. Stat3/Jak signalling and 2i inhibition alone cannot support the survival of <i>Dicer</i> <i>KO</i> cells	198
6.3.10. miRNA and epigenetic regulation of ground state pluripotency	198
6.4 Discussion	200

Chapter 7 The pleiotropic effects of retinol and its derivatives on pluripotency

7.1 Introduction	205
7.1.1. Retinoic acid and development	205
7.1.2. Retinoid metabolism	205
7.1.3 Mechanism of retinoic acid signalling	206
7.1.4. Retinoic acid, embryogenesis and ESC differentiation	207
7.1.5. Retinoids and pluripotency	208
7.1.6. Non-canonical retinoid signalling	208

7.2 Experimental aims.....	209
7.3 Results.....	210
7.3.1. Vitamin A enhances DNA demethylation during epigenetic reprogramming to the naive ground state	210
7.3.2. Retinol is the active metabolite	211
7.3.3. Expression of the ATRA signalling pathway is not upregulated in 2i with vitamin A	214
7.3.4. 2i plus Vitamin A activates PI3K and Akt signalling	216
7.3.5. Cell cycle progression is dysregulated in 2i minus Vitamin A conditions	219
7.3.6. Vitamin A and DNMT1 post translational modifications	221
7.4 Discussion	224

Chapter 8 Dissecting DNA demethylation and chromatin dynamics during PGC reprogramming

8.1 Introduction	228
8.1.1. Pioneers of germ cell biology	228
8.1.2. Primordial germ cell specification	228
8.1.3. Epigenetics of primordial germ cells	231
8.1.3.1 Migratory PGCs	231
8.1.3.2 Gonadal PGCs	232
8.1.3.3. The repair model of epigenetic reprogramming	233
8.1.3.4 De novo methylation in germ cells	234
8.2 Experimental aims	235
8.3 Results	236
8.3.1 Experimental system	236
8.3.2 <i>Dnmt1</i> KO PGCs fail to develop into mature male germ cells	237
8.3.3 DNMT1 is depleted prior to entry into the genital ridges	240
8.3.4. Developing ex vivo models of germline <i>Dnmt1</i> knockout	242
8.3.5. Derivation and characterisation of <i>Dnmt1</i> f/f mESCs	244
8.4 Discussion	246
Conclusion	250
Bibliography	255

Frequently used abbreviations

AID	Activation induced cytidine deaminase
APE1	Apurinic Apyrimidinic endonuclease 1
ATRA	All-trans retinoic acid
BER	Base excision DNA repair
BM Φ	Bone marrow derived macrophage
BMP	Bone morphogenic protein
BSA	Bovine serum albumin
CID	Collision induced dissociation
CGI	CpG island
CpG	5'-C-phosphate-G-3'
DAPI	4',6-diamidino-2-phenylindole
dCTP	Deoxycytidine phosphate
DMEM	Dulbecco's modified eagle's medium
DNMT	DNA methyltransferase
DKO	Double knockout
dTTP	Deoxythymidine triphosphate
E	Embryonic day
EC	Embryocarcinoma cells
EDTA	Ethylenediaminetetraacetic acid
EG	Embryonic germ cell
EdU	5-Ethynyl-2'-deoxyuridine
EMT	Epithelial to mesenchymal transition

EMV	Electron multiplier current
ERK	Extracellular signal regulated kinase
ESI	Electrospray ionisation
FACS	Fluorescent activated cell sorting
FCS	Foetal calf serum
Fgf	Fibroblast growth factor
FLM	FCS, LIF & MEF culture conditions
FOX	Forkhead box
GADD45	Growth arrest and DNA damage-inducible 45
GATA	GATA binding protein
GSK3 β	Glycogen synthase kinase 3 beta
HDAC	Histone deacetylase
hESC	Human embryonic stem cell
HPLC	High performance liquid chromatography
HTF	HpaII tiny fragment
HP1	Heterochromatin protein 1
ICM	Inner cell mass
KMT	Lysine Methyltransferase
LC-MS	Liquid chromatography mass spectroscopy
LIF	Leukaemia inhibitory factor
LOD	Limit of detection
LOQ	Limit of quantification
MAGE	Melanoma antigen gene
MBD	Methyl binding domain
M-CSF	Macrophage colony stimulating factor
mESC	Mouse embryonic stem cell

MEF	Mouse embryonic fibroblast
miRNA	micro RNA
MMR	Mismatch DNA repair
MRM	Multiple reaction monitoring
MTase	Methyltransferase
m/z	Mass to charge ratio
ncRNA	Non-coding RNA
OSKM	OCT4, SOX2, KLF4 & MYC
Parp	Poly(ADP-ribose) polymerase
PcG	Polycomb group protein
PGC	Primordial germ cell
PCNA	Proliferating cell nuclear antigen
QQQ	Triple quadrupole mass spectrometer
RRBS	Reduced representation bisulphite sequencing
RdDM	RNA-directed DNA methylation
ROS1	Repressor of silencing 1
SNR	Signal-to-noise ratio
SSEA1	Stage specific embryonic antigen 1
STAT	Signal transducer and activator of transcription
TE	Trophectoderm
TET	Ten Eleven Translocation protein
TDG	Thymine DNA glycosylase
Uhrf1	Ubiquitin like with PHD and RING finger domains 1
UMP	Uridine monophosphate
UTR	Untranslated region
VE	Visceral endoderm

XRCC1	X-ray repair cross-complementing protein 1
5CaC	5-carboxylcytosine
5fC	5-formylcytosine
5hmdC	5-hydroxymethyl-2' deoxycytosine
5mdC	5-methyl-2'deoxyctidine

List of figures

Chapter 1: An introduction to DNA methylation and the epigenetic regulation of pluripotency and embryonic development

Fig. 1.1 CpG island promoters

Fig. 1.2 CGI promoter methylation and transcriptional activation

Fig. 1.3 Mechanism of DNA methylation by MTase and SAM

Fig. 1.4 Overview of TET-catalysed 5mdC oxidation

Fig. 1.5 5'hydroxymethylcytosine and passive DNA demethylation

Fig. 1.6 5hmdC and active DNA demethylation

Fig. 1.7 DNA methylation dynamics during embryonic development

Fig. 1.8 Schematic of preimplantation development and the regulation of cell fate decisions

Fig. 1.9 LIF induces the JAK/STAT signalling cascade

Fig. 1.10 Modulation of self-renewal and differentiation by GSK3 β and ERK inhibition

Chapter 3: Quantification of global DNA modifications by LC-MS

Fig. 3.1 Electrospray ionisation

Fig. 3.2 Schematic of the common scanning modes used during QQQ mass spectrometry

Fig. 3.3 Sample dilution and the reduction of ion suppression

Fig. 3.4 Ion suppression and the effect of DNA extraction methods

Fig. 3.5 Glycerol is a major source of ion suppression

Fig. 3.6 Quantification of 5mdC in small cell numbers

Fig. 3.7 Cytological fixation techniques and their effect on LC-MS ion suppression

Chapter 4: *In vivo* and *ex vivo* DNA methylation dynamics

Fig. 4.1 Stable isotope labelling of pyrimidines and 5-methylcytosine

Fig. 4.2 Pathways of pyrimidine biosynthesis

Fig. 4.3 Validation of LC-MS method to detect individual isotopomers

Fig. 4.4 *In vivo* ^{13}C -d₃ L-methionine labelling of adult cortical neurons

Fig. 4.5 $^{15}\text{N}_2$ Uridine concentration on the incorporation of $^{15}\text{N}_2$ -dCTP and $^{15}\text{N}_2$ -dTTP

Fig. 4.6 Stable isotope labelling of *Tet1* and *Tet2* double KO (DKO) mESCs

Chapter 5: Regulation of global DNA methylation in mESCs upon transition to 2i media

Fig. 5.1 The dynamics of DNA demethylation in migratory PGCs are similar to those induced in mESCs upon transition to 2i

Fig. 5.2 Cell lines and the culture conditions used in this study

Fig. 5.3 LC-MS quantification of 5mdC in mESCs upon transition to 2i

Fig. 5.4 LC-MS quantification of 5hmdC in mESCs upon transition to 2i

Fig. 5.5 Transcriptional downregulation of *de novo* Dnmts occurs rapidly upon transition of mESCs to 2i

Fig. 5.6 Culture-induced changes to the cell cycle and proliferation dynamics

Fig. 5.7 Tet expression and TET-driven DNA demethylation

Fig. 5.8 Prdm14 expression and DNA demethylation in 2i

Chapter 6: miRNAs as regulators of naive and ground state pluripotency

Fig. 6.1 Dicer is differentially expressed in FLM and 2i culture conditions

Fig. 6.2 LC-MS quantification of 5mdC in *Dicer KO* and WT cells

Fig. 6.3 DICER is necessary for *Dnmt3a* isoform 2 downregulation in 2i

Fig. 6.4 miRNAs significantly up- or down-regulated in 2i

Fig. 6.5 Alternative means of miRNA-mediated regulation

Fig. 6.6 miRNA regulation of *Dnmt3a* isoform 2 expression

Fig. 6.7 Expression of miRNAs targeting *Dnmt3a* isoform 2 transcriptional regulators

Fig. 6.8 Expression of *Dnmt3a* isoform 2 transcriptional regulators in FCS and 2i

Fig. 6.9 Exploratory analysis of *Dicer KO* RNA-Seq data

Fig. 6.10 Differential gene expression in *Dicer KO* cells cultured in 2i or FLM growth conditions

Fig. 6.11 miRNAs and the control of the pluripotency transcriptional network

Fig. 6.12 miRNAs and the epithelial-mesenchymal transition

Fig. 6.13 Expression of epigenetic writers and readers mediated by post-transcriptional regulation

Chapter 7: The pleiotropic effects of retinol and its derivatives on pluripotency

Fig. 7.1 Retinoic acid signalling

Fig. 7.2 Effect of vitamin A on cell morphology and DNA modifications

Fig. 7.3 LC-MS analysis of the effects of retinol (Vit A) on epigenetic reprogramming

Fig. 7.4 Effect of vitamin A on the transcriptional expression of DNA methyltransferases and their effectors

Fig. 7.5 The presence of vitamin A causes an upregulation of the pluripotency transcriptional circuitry

Fig. 7.6 Retinol (Vit A) does not activate the retinoic acid signalling pathway

Fig. 7.7 Akt signalling in the presence of vitamin A

Fig. 7.8 The absence of vitamin A causes defects in cell cycle progression

Fig. 7.9 Culture induced changes in UHRF1 protein expression

Fig. 7.10 Colocalisation of DNMT1 and EdU at replicating foci

Chapter 8: Dissecting DNA demethylation and chromatin dynamics during the epigenetic reprogramming of primordial germ cells

Fig. 8.1 Stages of murine primordial germ cell migration

Fig. 8.2 DNA demethylation dynamics during PGC development

Fig. 8.3 Cre-mediated *Dnmt1* deletion

Fig. 8.4 Phenotype of germline *Dnmt1* deletion in adult gonads

Fig. 8.5 DNMT1 is depleted in E10.5 PGCs

Fig. 8.6 Loss of germ cells in male E14.5 gonads

Fig. 8.7 Derivation of *Dnmt1* *F/F* mESCs for ex vivo *Dnmt1* deletion in PGCLCs

List of Tables

Table 2.1 Mouse genotyping primers and PCR protocols

Table 2.2 Mouse embryonic stem cell lines

Table 2.3 Antibodies used against BM Φ cell surface proteins

Table 2.4 Immunostaining of cryosections

Table 2.5 Preparation of LC-MS standard curve

Table 2.6 RT-qPCR primer sequences

Table 2.7 Primary antibodies used for Western blots

Table 3.1 Summary of flow-dependent LC-MS parameters

Table 3.2 Summary of compound dependent parameters

Table 3.3 Properties of the HPLC columns tested

Table 4.1 Summary of additional compound dependent parameters

Chapter 1

An introduction to DNA methylation and the epigenetic regulation of pluripotency and embryonic development

1.1 Understanding biological complexity in the post-genomic era

1.1.1 The hopes of whole genome sequencing

The Human Genome Project was launched with the hope that, by cracking the code of life, biology's Enigma, we could better understand what makes us human. The sequencing of the human genome is arguably one of our greatest scientific achievements. Invention and rapid advances in recombinant and sequencing technologies enabled the publication of the first working draft sequence of the Human Genome Project in 2001; fewer than 50 years after James Watson and Francis Crick published their results on the molecular structure of nucleic acids (Lander et al., 2001)(Watson & Crick, 1953). Remarkably, the number of genes that make up the protein coding portion of the human genome is comparable to that found in lower organisms; 20,000 rather than the initial estimates of 50,000 to 140,000. However these genes account for only 1.5 percent of the entire genome; the rest likely to include large regions of regulatory DNA and repetitive elements (Venter et al., 2001). Upon its publication, Francis Collins, director of the National Human Genome Research Institute likened the genome to a multipurpose history book, instruction manual and medical encyclopedia; providing scientists with a narrative of our species' evolution, a detailed blueprint of the human cell and pinpointing disease genes and drug targets. However, despite its many accomplishments, the Human Genome Project has created many more questions than answers. In particular, the post genomic era has focused on understanding how genetic information is spatially and temporally interpreted by somatic cells.

1.1.2 Epigenetic regulation increases the genome's information content

Significant incongruities exist between biological complexity, genome size and the number of classically defined protein coding genes. Eukaryotic genome size covers a 100,000-fold range, in which the single celled *Amoeba dubia* is among the largest (Pellicer et al., 2010). Wild variation is also observed between closely related clades; amphibians for example show a 120-fold range in genome size from frogs to salamanders (Licht & Lowcock, 1991). This enigma is known as the C-value paradox (the C-value being the haploid nuclear DNA content)

(Cavalier-Smith, 1985). Furthermore, complexity is not defined by the number of genes; the G value paradox (Hahn & Wray, 2002). Biological complexity instead correlates with the number of regulatory processes able to 'fine-tune' the interpretation of the underlying genetic sequence (Szathmary et al., 2001)(Valentine, 2000).

The several hundred distinct cell types within the human body are genetically homogeneous; bar a few notable exceptions (including lymphocytes and post-meiotic germ cells) which have undergone recombination. Only a fraction of the genome is expressed within a given cell. Whilst the expression of housekeeping genes shows very little variation across cell types, many other genes are under restrictive cell-type, spatial and temporal control. Faithful cell-type specific interpretation of the genome is necessary to produce an array of functionally and structurally heterogeneous cell types. The dynamic changes to the transcriptional network during cell differentiation rely on altered, combinatorial binding of transcription factors to gene-proximal promoters and distal regulatory regions (Zinzen et al, 2009)(Farnham, 2009). The determinants of transcription factor binding include direct and indirect (cofactor) sequence preference and the local chromatin context. Large scale, genome-wide profiling of binding sites, such as the ENCODE consortium, have shown that cell-type specific usage of regulatory elements is typically associated with a change in the surrounding chromatin (Arvey et al., 2012). These non-genetic (epigenetic) modifications alter the ability of transcription factors to bind to their cognate DNA motifs, and thus define the context or landscape of cell-selective transcriptional programmes.

Epigenetics can broadly be described as a bridge between genotype and phenotype. The Greek preposition 'epi' can signify a superimposition or can be translated to 'on', 'before', 'after' or 'close to'. Therefore, literally epigenetics translates to 'in addition' or 'outside' conventional genetics. In other words, gene function is the sum of not only genetic information, but also how this information is interpreted in a given context. Epigenetic states can be stably inherited through mitosis. Epigenetic regulation occurs on three major levels; DNA, histones, and nucleosomes.

1.1.3 Gene expression is constrained by the chromatin landscape

The eukaryotic genome is assembled into a complex nucleoprotein structure known as chromatin. The basic repeating unit of chromatin, the nucleosome, consists of approximately 146 bp DNA wrapped around an octamer of histones. N-terminal histone tails protruding from the nucleosome are known to be subjected to an ever-increasing number of reversible post-translational modifications, including acetylation, phosphorylation, methylation, ribosylation and sumoylation (Murray, 1964)(Kleinsmith et al., 1966)(Vidali et al., 1968)(Shilo & Eisenman, 2003). Histone modifications affect inter-nucleosomal interactions and recruit proteins and complexes with specific enzymatic activities. To this end, different modifications are associated with the altered activity of promoters and enhancers.

Histone modifications associated with transcriptionally active loci recruit proteins required to create or maintain an open chromatin state (euchromatin). Other proteins, including lysine methyltransferases (KMTs), heterochromatin protein 1 (HP1), histone deacetylases (HDAC), and polycomb proteins, create highly condensed, transcriptionally silenced chromatin (heterochromatin). Compact, higher order chromatin organisation sequesters a large proportion of the genome (approximately 70% in differentiated cells), creating a physical barrier to spurious transcription and mediating long-range, possibly interchromosomal interactions (Struhl, 1999)(Workman & Kingston, 1998).

However, since DNA replication requires access to the DNA, chromatin states must be dynamic and reversible. Similarly, changes to the transcriptional programme require histone displacement and chromatin remodelling (Chi et al., 2002)(Saladi & de la Serna, 2010). To facilitate the activation of selected loci, pioneer transcription factors, including the FOXA and FOXO families, access DNA target sites embedded within higher order chromatin; leading to the reorganisation of local chromatin structure, directly recruiting other chromatin modifiers and initiating the assembly of further regulatory factors.

1.2 DNA methylation and the regulation of cell fate

1.2.1 Defining DNA methylation

In addition to histone modifications, the epigenetic landscape is also defined by an abundance of DNA modifications. The most prevalent, DNA methylation (also known as methylcytosine, 5mdC), entails the covalent addition of a methyl (CH_3) moiety to the C5 position of cytosine. In mammalian somatic cells, DNA methylation is typically found within CpG dinucleotides. By altering the functional state without affecting the Watson-Crick base pairing, DNA methylation significantly increases the information content of the genome. DNA methylation states are also stably maintained through mitotic cell divisions. These latter points are important in defining 5mdC as an epigenetic mark rather than a DNA lesion.

1.2.2 The discovery of the fifth base

The discovery of 5mdC predates the structural determination of the DNA double helix. In 1948, the evidence for a link between 'genes' and the material of genetic inheritance was mounting, but as yet unproven. Rollin Hotchkiss, the distinguished biochemist and geneticist, developed a method for the chromatographic separation of purines and pyrimidines (Hotchkiss, 1948). In addition to completely resolving adenine, cytosine, guanosine, thymidine and uracil, a minor base was consistently found to migrate at a similar rate to thymidine (also known as methyluridine). This base was identified as 5mdC.

1.2.3 DNA methylation states are stably inherited

An early model for the biological role of DNA methylation in differentiation and development was independently proposed by Robin Holliday & John Pugh (1975) and by Arthur Riggs (1975). Since post-synthetic modifications were already known to impart regulatory effects on RNA and proteins, both groups argued that by modifying DNA-protein interactions, DNA methylation was also likely to affect gene expression. DNA methyltransferases (MTase) with a high affinity towards hemimethylated DNA had already

been isolated from bacteria. It was proposed that a similar MTase resided within mammalian cells. By binding to hemimethylated sites formed during DNA synthesis, a 'maintenance' DNA methyltransferase would enable symmetrical DNA methylation to be semiconservatively copied after DNA replication (Bird, 1978); thus providing a mechanism for the stable inheritance of DNA modifications.

1.2.4 The link between DNA methylation and transcriptional repression

The observation that Swiss 3T3 and C3H/10T12CL8 cells treated with a potent inhibitor of DNA methylation (5-azacytidine) developed phenotypes more typically associated with contractile striated myoblasts, adipocytes and chondrocytes forged a functional link between DNA methylation and gene repression (Taylor and Jones, 1979). Specifically, DNA methylation was shown to prevent spurious transcription, developmental reprogramming and lineage switching. Subsequent findings, including the observation that *in vitro* methylated DNA is transcriptionally inactive when transfected into *Xenopus* oocytes (Vardimon et al., 1982) or cultured mammalian cells (Vardimon et al., 1982)(Stein et al., 1982) added further credence to this theory. Unmethylated promoter-proximal CpG rich regions (CpG islands), initially known as the HpaII tiny fragment (HTF) fraction, were found to associate with the majority of vertebrate genes; thus physically linking DNA methylation and transcriptional regulation (Reviewed in Bird, 1987). The ability for DNA methylation states to regulate gene expression was confirmed by the *ex vivo* expression of hybrid methylated adenine phosphoribosyltransferase (*Aprt*) genes (Keshet, 1985). Specifically, gene body methylation had no effect of transcription whilst methylation of the 5' CpG island promoter dramatically reduced transcriptional activity.

1.2.5 DNA methylation and lineage restriction in post-implantation development

The differentiation of the epiblast into the three germ layers coincides with a generalised wave of *de novo* methylation (Howlett & Reik, 1991)(Kafri et al., 1992). Subsequently, changes in DNA methylation are most often gene or tissue-specific. In both instances, DNA methylation is integral for the specification of cell fate and lineage restriction. This conjecture is irrevocably proven by a multitude of observations; including the deleterious developmental consequences of *in vivo* DNA methyltransferase (DNMT) deletion (Li et al., 1992),

the dynamic genome-wide DNA methylation changes seen during gametogenesis, cell type specific methylomes (Deaton et al., 2011)(Laurent et al., 2010)(Doi et al., 2009), the repressive effect of *in vitro* methylated promoters and distal regulatory elements (Boyes & Bird, 1991), the regulation of enhancer activity and transcription factor binding (Lister et al., 2009)(Sérandour et al., 2011)(Schmidl et al., 2009), gene dosage compensation by X-chromosome inactivation, parent-of-origin allelic expression and the contribution of altered DNA methylation states to human diseases including cancer, immunological and neurological disorders (Feinberg & Vogelstein, 1983)(Gama-Sosa et al., 1983)(Esteller, 2008)(Dittrich et al., 1992)(Jin et al., 2008).

1.3 Structure of the DNA methylome

1.3.1 Phylogenetic variation in DNA methylation

Developmental transcriptional changes in many metazoans, including *Caenorhabditis elegans* and dipteran insects, are not dependant on DNA methylation-mediated transcriptional silencing (Simpson et al., 1986)(Rae & Steele, 1979). These organisms, which have very low or undetectable levels of 5mdC, rely instead on polycomb-group (PcG) and trithorax group (TrxG) proteins, miRNA and piRNA-dependent transcriptional silencing and histone modifications to maintain heritable transcriptional states (Birchler et al., 2000)(Ingham, 1985)(Gowher et al., 2000). In many other clades, DNA methylation is obligatory but nonetheless phylogenetically variable. In plants and invertebrates, mosaic DNA methylation patterns regulate specific genomic elements including mobile genetic elements (self-contained genomic units capable of proliferating within their host genome) and actively transcribed sequences (Bird et al., 1979)(Montero et al., 1992). At the extreme end of the spectrum, extensive DNA methylation is the default epigenetic state for vertebrates. The genetic ablation of DNA methyltransferases results in developmental arrest during early embryonic development and the widespread derepression of ectopic gene expression and transcriptional activation of transposable elements (Panning et al., 1996)(Stancheva et al., 2001)(Jackson-Grusby et al., 2001)(Walsh et al., 1998).

1.3.2 CpG island promoters

Pervasive DNA methylation does come at a price. 5mdC is mutagenic and sporadic deamination of 5mdC to thymine generates G:T mismatches and severe genome-wide CG depletion (21% of the expected proportion in the human genome)(Lander et al., 2001). Short interspersed sequences containing atypically high CpG content are an exception to this rule. Identified by methylation sensitive (HpaII) restriction endonucleases, these CpG islands (CGI) are typically unmethylated, and therefore do not have the same mutagenic properties as methylated CpG dinucleotides (Cooper et al., 1983)(Bird et al., 1985). More recent technological advances enabling the mapping of DNA methylation sites at an unprecedented resolution have shown that CGIs are associated with 60-70% of human genes; making them the most common class of promoter in the human genome (Saxonov et al., 2006)(Illingworth & Bird., 2009). Expression of virtually all ubiquitous housekeeping genes and approximately 40% tissue specific and developmentally regulated genes are controlled by CGI promoters (Larsen et al., 1992)(Zhu et al., 2008). Many other 'orphan' CGIs lie within or between transcriptional units (Illingworth et al., 2010). Evidence, including the colocalization of active marks of transcription (H3K4me3) and sequencing of the 5' end of capped transcripts (CAGE), suggests that these CGIs represent novel promoters; potentially being alternative promoters for nearby genes or for the expression of non-coding RNA (ncRNA)(Shiraki et al., 2003).

By and large, CGI-promoters remain unmethylated regardless of the transcriptional state. How CGI-promoters escape DNA methylation is an active area of research. Alternative models include the occlusion of DNA methyltransferases by transcription factor binding (Fig. 1.1b) or localised active DNA demethylation (Fig. 1.1c) (Feldmann et al., 2013)(Hsieh et al., 1999)(Lin et al., 2001)(Wu & Zhang, 2010). In the absence of DNA methylation, CGI-linked genes can be repressed by polycomb group proteins and H3K27me3 (Fig. 1.1a).

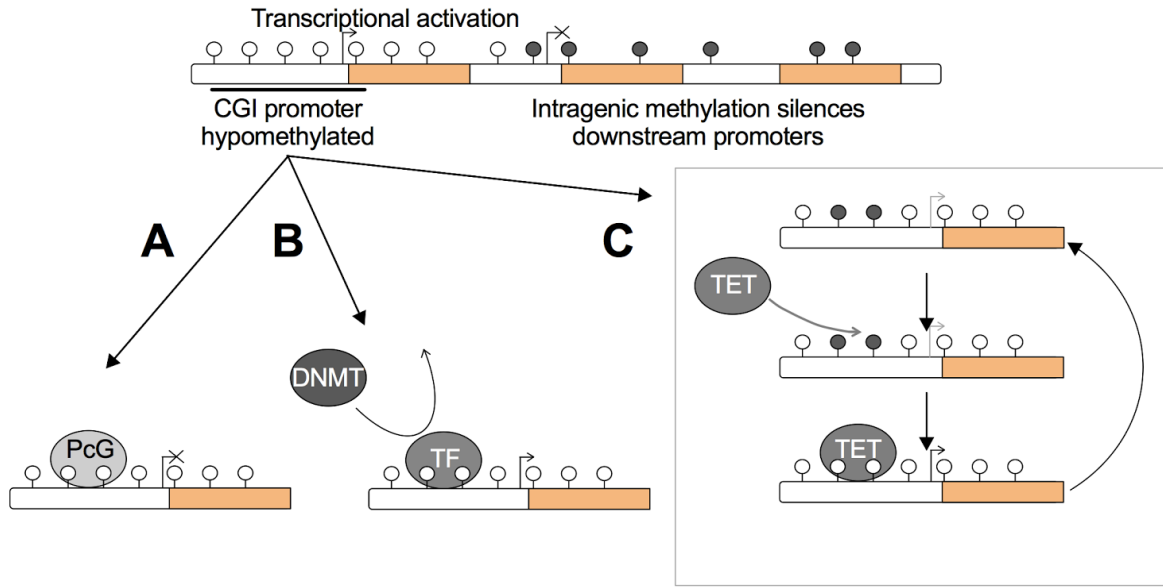


Figure 1.1 CpG island promoters

CpG islands (CGIs) are typically hypomethylated.

(a) The majority of mammalian promoters are associated with a CGI. Most CGI-promoters are hypomethylated regardless of their transcriptional state. Activation or repression of gene expression is likely to instead be regulated by histone modifications and polycomb group proteins.

(b) How CGI-promoters remain hypomethylated is not clear. One model proposes that DNA hypomethylation is maintained by the competition for binding sites between DNA methyltransferases, methyl-binding proteins and transcription factors.

(c) Alternatively, following sporadic, aberrant DNA methylation, promoters may be actively demethylated (i.e. by TET-mediated hydroxylation).

A small but significant proportion of promoter-CGIs are, however, known to acquire methylation during development; some being associated with X-chromosome inactivation in females and parent of origin allelic imprinting, whilst others are differentially methylated between different cell types. Extensive evidence suggests that CGI-promoter methylation plays a role in transcriptional repression. However, there are also many genes which display a remarkably poor correlation between promoter methylation and transcriptional status (Eckhardt et al., 2006). These disparities are potentially explained by the use of alternative promoters; hypermethylation of a primary CGI-promoter may lead instead to initiation from a downstream intragenic promoter (Fig. 1.2). As an example, the cancer testis (CT) antigens belonging to the Melanoma-associated antigen (MAGE) family show differential, lineage specific methylation; being hypermethylated in somatic tissues and hypomethylated in the germline. DNA hypomethylation in many cancers results in the ectopic expression of these MAGE proteins, suggesting that DNA methylation may be the primary mode of gene repression. Other studies have shown that constitutively silenced CGI promoters possess variable degrees of

DNA methylation, indicating that DNA methylation is not necessary for transcriptional silencing. In these instances, *de novo* DNA methylation may be stochastic and the result of reduced transcription factor binding, enabling occlusion of DNMTs from the promoter-CGI.

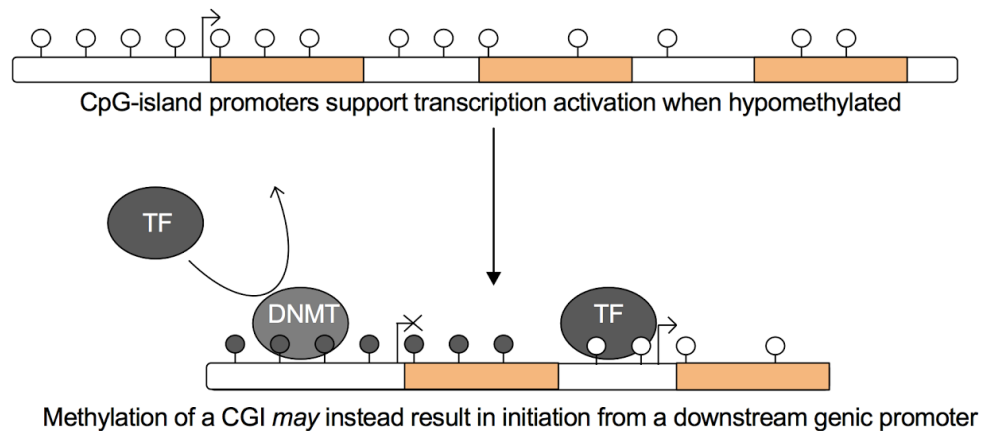


Figure 1.2 CpG island promoter methylation and transcriptional activation

A small proportion of CGI-promoters acquire DNA methylation during the course of development. Although there is a strong correlation between CGI promoter methylation and transcriptional repression, it is not clear whether DNA methylation is causative. Several other studies have shown that DNA methylation alone cannot silence gene expression. It has been proposed that DNA methylation of a CGI promoter results in the use of alternative downstream promoters. Thus providing an explanation for why the methylation of a promoter CGI does not necessarily silence gene expression.

1.3.3 The paradox of gene body methylation

High levels of DNA methylation are also found in gene bodies of actively transcribed genes. Rather than being associated with gene repression, for instance by preventing transcriptional elongation, the level of gene body methylation positively correlates with the level of gene expression (Wolf et al., 1984)(Hellman & Chess, 2007)(Cokus et al., 2008)(Lister et al., 2009). Indeed, the remethylation of genomic DNA in cancer cells treated with DNA methyltransferase inhibitors occurs more rapidly at actively transcribed genes and is required for their correct expression (Yang et al., 2014).

Although evolutionarily conserved, genic DNA methylation is poorly understood. Studies using *Saccharomyces cerevisiae*, which lacks DNA methylation, showed that the deposition of H3K36me and subsequent recruitment

of histone deacetylases to actively transcribed sequences creates a densely packed chromatin structure and eliminates spurious transcription from genic cryptic start sites (Carrozza et al., 2005). The need to stabilise chromatin structure caused by elongating RNA polymerase is even more dire for multicellular organisms which have substantially longer genes; many with intragenic promoters and retroviral elements. Therefore, DNA methylation has been proposed to suppress intragenic promoters (Maunakea et al., 2010), retroviral elements and other functional elements which could otherwise reduced the transcriptional efficiency and Pol II processivity. Finally, single molecule bisulphite sequencing has identified a subtle enrichment of CpG methylation in exons compared to their flanking introns; pointing to an emerging role for intragenic DNA methylation in both constitutive and alternative splicing (Hodges et al., 2009). DNA methylation may fine-tune alternative splicing by reducing the rate of Pol II elongation or by recruiting splicing factors. The leukocyte common antigen CD45, for instance, is alternatively spliced during lymphocyte differentiation. Binding of CTCF to exon 5 reduces RNA pol II processivity, leading to increased inclusion of this alternative exon in the mature, spliced transcript. DNA methylation of exon 5, however, prevents CTCF binding and subsequent RNA pol II pausing, thus reducing the likelihood of exon inclusion (Shukla et al., 2011)(Ong et al., 2014). Therefore, genic DNA methylation may support the development of complex multicellular organisms by increasing the transcriptome and proteome diversity.

1.3.4 Non-CpG methylation

The vast majority of research into DNA methylation has focused on CpG palindromes. However, methylation of cytosines in alternative sequence contexts is known to be prevalent in plants and some fungi. The existence of methylated non-CpG dyads was initially confirmed in *Arabidopsis* (Cokus et al., 2008). Subsequently, non-CpG methylation was discovered in mammals, but for many years was thought to exclusively be a feature of pluripotent cells. In human embryonic stem cells (hESCs), for example, non-CpG methylation is abundant; accounting for approximately 25% of all methylated cytosines (Lister et al., 2009). Non-CpG methylation is enriched in gene bodies. Core promoters, transcription start sites and regions with high density of known protein binding sites however have low levels of non-CpG methylation. More recent studies, however have provided evidence that non-CpG methylation does persist through organogenesis, with non-CpG methylation

being found in the adult mammalian brain, skeletal muscle and hematopoietic stem cells (Guo et al., 2014)(Yan et al., 2011). The function, if any, of non-CpG methylation is currently unknown.

1.4 DNA methyltransferases

1.4.1 Maintenance methylation

DNA methyltransferase 1 (*Dnmt1*) is ubiquitously and highly expressed in replicating somatic cells. Due to its preference for hemimethylated palindromic CpGs, DNMT1 is known as a maintenance methyltransferase (Bestor & Ingram, 1983)(Hermann et al., 2004). In the immediate aftermath of DNA replication CpGs in the newly synthesised DNA will be unmethylated. DNMT1 associated with the replication machinery recognises the asymmetry of the hemimethylated CpGs prominently displayed within the major groove of the DNA double helix. DNA methyltransferase activity is therefore directed against the unmethylated cytosine within the CpG dyad such that DNA methylation states are faithfully restored following DNA synthesis.

Bioinformatic analysis suggests that *Dnmt1* has evolved by the fusion of three ancestral genes (Margot et al., 2000). At its simplest, Dnmt1 consists of a N-terminal regulatory region connected to a C-terminal catalytic region via seven lysyl glycyl dipeptide repeats. The regulatory region, contains a CXXC zinc finger domain, a PCNA-binding motif, two bromo adjacent homology domains (BAH1/2) and a targeting sequence (TS). In combination, these motifs aid the binding of DNMT1 to unmethylated CpG and mediate the interaction between DNMT1 and the replication machinery. Despite containing all essential catalytic motifs, the C-terminal catalytic domain is not sufficient to catalyse the transfer of a methyl moiety alone. Two crystal structures have alluded to the allosteric activation of DNMT1 via the regulatory domains (Seyda et al., 2011)(Takeshita et al., 2011)(Song et al., 2011). Both structures show a similar interaction between the C-terminal and N-terminal domain. Specifically, the targeting sequence was found to be deeply inserted within the catalytic pocket, preventing any interaction between DNMT1 and unmethylated DNA. Activation can only occur once this autoinhibitory mechanism is abolished by displacement of the TS domain.

DNMT1 targeting, specificity and activity are influenced by a large number of protein interacting factors (Reviewed in Qin et al., 2011). Of particular interest are the methyl CpG binding proteins including MeCP2, MBD2, MBD3, and the UHRF family. MeCP2, MBD2 and MBD3 interact with histone deacetylase (HDAC) 1 and 2 (Jones et al., 1998). Binding of DNMT1 to methyl CpG binding proteins therefore links DNA methylation and histone deacetylase activity. Both modifications contribute to the formation of transcriptionally inactive heterochromatic regions. UHRF1 on the other hand is a crucial DNMT1 cofactor. Quantification of 5mdC by Southern blot and genomic bisulphite sequencing suggested that genetic ablation of *Uhrf1* in mouse embryonic stem cells (mESCs) leads to hypomethylation similar to that seen in *Dnmt1* ^{-/-} mESCs (Bostick et al., 2007)(Sharif et al., 2007). Additionally, the kinetics of DNA demethylation upon loss of either *Dnmt1* or *Uhrf1* in mESC are comparable (Von Meyenn et al., 2016). Throughout S-phase, DNMT1 and UHRF1 colocalize. The preferential binding to hemimethylated DNA is facilitated by UHRF1's SET and RING associated (SRA) domain; suggesting that Dnmt1 may be recruited to replication foci by UHRF1, rather than binding its substrate directly (Avvakumov et al., 2008)(Hashimoto et al., 2008)(Arita et al., 2008).

Other factors that either increase the processivity of DNMT1, or regulate its recruitment include chromatin remodelling ATPases, DNMT1-associated protein (DMP1) and direct interactions with transcription factors (SP1, SP3 and STAT3) (Rountree et al., 2000). Finally post-translational modifications have been shown to affect DNMT1 localization and protein stability. In all, the maintenance of DNA methylation relies on the coordination of multiple different protein activities.

1.4.2 *De novo* methylation

For a decade, DNMT1 was thought to be the sole mammalian DNA methyltransferase. Several observations, however, suggested that this may not be true. Firstly, genetic knockout of *Dnmt1* in mESCs only produced a partially demethylated genome (Lei et al., 1996). Following transfection of these cells with retroviral constructs, the exogenous DNA underwent *de novo* methylation; indicating that at least one other DNA methyltransferase was active. Analysis of expression sequence tags (ESTs) led to the identification and cloning of two potential DNA methyltransferases based on their sequence similarity to conserved motifs in bacterial

C5-methyltransferases: *Dnmt3a* and *Dnmt3b* (Okano et al., 1998)(Yoder & Bestor, 1997). Overexpression of DNMT3A and DNMT3B proteins in a baculovirus system confirmed that both proteins had intrinsic DNA methyltransferase activity.

DNMT3A and DNMT3B expression levels correlate with periods of high *de novo* DNA methyltransferase activity, including during blastocyst and germ cell development (Okano et al., 1999). Hence, they are known as *de novo* methyltransferases. During differentiation, *Dnmt3a* and *Dnmt3b* undergo tissue specific transcriptional repression. Loss of either *de novo* methyltransferase is lethal (Okano et al., 1999). *Dnmt3b* KO mice die around midgestation as a result of a myriad of developmental defects including rostral tube defects, liver hypertrophy and hemorrhaging. *Dnmt3a* KO mice meanwhile typically die within 4 weeks of birth showing symptoms of gastrointestinal dysfunction.

The molecular structure of DNMT3A and DNMT3B enzymes is clearly distinct from the structure of the maintenance methyltransferase DNMT1. Instead, clear sequence and structural similarities between the catalytic domain of the *de novo* Dnmts and the bacterial C5-methyltransferase M.HhaI suggest a common ancestor (Jia et al., 2007)(Cheng et al., 2008). In particular, key residues including the catalytic cysteine in motif IV and the glutamic acid in motif VI are conserved (Fig. 1.3). Mutation of these residues has confirmed that they are essential for *in vivo* and *in vitro* methyltransferase activity (Gowher et al., 2006). These parallels suggest that DNMT3A and DNMT3B may use a similar catalytic mechanism to bacterial M.HhaI.

The primary physiological function of the *de novo* methyltransferases is to establish DNA methylation patterns during cell fate determination and differentiation. This is achieved with the aid of their cofactor, Dnmt3-like (DNMT3L) methyltransferase. DNMT3L has a truncated pseudo catalytic domain lacking the key residues and is therefore catalytically inactive (Aapola et al., 2000). *Dnmt3l* *-/-* animals are viable and typically asymptomatic. However, offspring from a *Dnmt3l* *-/-* female and wild type male are known to die during embryonic development as a result of severe developmental delays, neural tube defects and placental abnormalities (Bourc'his et al., 2001). These phenotypes are most likely caused by the aberrant expression of multiple

imprinted loci: including *Snrpn*, *Igf2r*, *Peg1*, *Peg3* and *Cdkn1* (Bourc'his et al., 2001)(Hata et al., 2002), suggesting that although DNMT3L may not be required for post-zygotic embryonic development, its presence is required for the correct establishment of maternal imprints during oogenesis. Similarly, *Dnmt3l* $-/-$ males are infertile, with reduced testes size associated with complete azoospermia. Ultimately, this phenotype is caused by transcriptional derepression of repetitive elements and the large scale, global failure of chromosomal synapsis during the leptotene and zygotene stages of meiosis (Bourc'his & Bestor, 2004). Coimmunoprecipitation indicates that DNMT3L directly interacts with DNMT3A and DNMT3B. This protein interaction appears to take place between the catalytic and pseudo-catalytic domains. Consequently, *de novo* Dnmt activity is stimulated by up to 20-fold and exhibits increased affinity towards the cofactor S-adenosyl methionine (also known as AdoMet, or SAM) (Gowher et al., 2005).

1.5 The tumultuous search for a putative demethylase

1.5.1 Dynamic and static DNA methylation states

Lineage restriction, cell fate determination and ultimately organogenesis rely on the non-genetic inheritance of DNA methylation through mitotic divisions. Symmetrically methylated CpG dinucleotides are copied semiconservatively from parent to daughter strand during DNA replication, ensuring that chromatin is permissive specifically to the transcriptional programme of the mother cell.

Although typically regarded as stable, any model of DNA methylation must also accommodate the highly orchestrated switching of gene expression programmes that occurs during embryonic development. Furthermore, DNA methylation patterns are invariably disturbed by DNA replication, repair and transcription, resulting in substantial, dynamic turnover, at least in some genomic contexts. In the most extreme example of epigenetic plasticity, genome-wide DNA methylation erasure occurs during periods of global reprogramming in the zygote and development of primordial germ cells (PGCs)(Refer to Chapter 8 for a description of epigenetic reprogramming in PGCs).

1.5.2 Passive DNA demethylation

Passive DNA demethylation describes the replication-dependent dilution of 5mdC due to the inhibition of DNA methyltransferase activity, nuclear exclusion or the down-regulation of the DNA methylation machinery. In other words, anything that prevents the faithful recruitment and enzymatic activity of DNMT1 on a hemimethylated DNA template would lead to a progressive loss of DNA methylation over successive cell divisions. There are several developmentally significant examples of passive DNA demethylation, including the dilution of 5mdC from the maternally derived genome during the cleavage stage of preimplantation development (Mayer et al., 2000). This process is likely to be driven by the exclusion of the oocyte-specific DNMT1 α from the nucleus of the early embryo (Mertineit et al., 1998)(Cardoso et al., 1999).

1.5.3 Active DNA demethylation in plants

Active DNA demethylation relies on the enzymatic, replication-independent removal of 5mdC or its derivatives. Whilst the molecular basis for active DNA demethylation in mammals is highly controversial, evidence for the same process in plants is concrete.

In plants, RNA-directed DNA methylation (RdDM) targets transposons and repetitive elements distributed throughout plant genomes (Jones et al., 1998)(Wassenegger et al., 1994) and is important for suppression of mobile elements in species with larger genomes, including rice and maize. Repressor of Silencing 1 (ROS), a bifunctional DNA glycosylase lyase (Gong et al., 2002)(Morales-Ruiz et al., 2006) was discovered in genetic screens for mutants with promoter hypermethylation and transgene silencing effects. ROS1 and its paralogs (DME, DML2 and DML3) actively remove 5mdC via base excision DNA repair (BER). *In vivo*, mutations in these genes cause increased DNA methylation in all sequence contexts at specific genomic loci; suggesting locus specific targeting (Lister, 2008)(Gong, 2002).

1.5.4 The early search for a mammalian demethylase

Mammalian protease sensitive DNA demethylase activity, detected as the ability to excise tritiated methyl groups from DNA, was initially reported in the nuclear extracts of MEL cells (Gjerset and Martin, 1982). Later experiments suggested that the elusive demethylase may actually be ribozyme-like (Weiss, 1996). However, these early experiments were inconclusive and fraught with technical concerns (Swisher, 1998). Nonetheless, despite the controversy surrounding the initial searches for a putative mammalian DNA demethylase, the presence of active DNA demethylation in mammalian systems is now irrefutable.

Firstly, it may be argued that, since DNA methyltransferases are abundantly expressed in non-dividing cells, the absence of active DNA demethylation would cause the progressive accumulation of DNA methylation over time. This is known not to be the case (Numata et al., 2012). Additionally, the activity dependent reactivation of genes including *Nr2b* and *Bdnf* by DNA demethylation has been observed in neurons. Furthermore, pharmacological inhibition or deletion of DNA methyltransferases in postmitotic neurons results in decreased 5mdC levels (Feng et al., 2010)(Miller et al., 2007). Finally, cytosine analogs including 5'Azacytidine have been shown to incorporate into the DNA of non-dividing cells at rates which, although low, were high enough to physically trap DNA methyltransferases (Yamagata et al., 2012). However, no consensus putative mammalian DNA demethylase has been identified and the mechanisms thus remain elusive.

1.5.5 DNA demethylation by direct 5mdC excision

The first and conceptually simplest mechanism to be explored involved the direct removal of the methyl moiety. However, so far, efforts to identify a mammalian ROS1/Demeter family ortholog capable of directly excising 5mdC have proven fruitless.

Methyl-CpG binding domain (MBD) protein 2 and MBD4 were identified in a database search for proteins with a MBD resembling that of the transcriptional repressor MECP2 (Hendrich & Bird, 1998) and independently in yeast screens for proteins involved in strand discrimination (Bellacosa et al., 1999). *Bona fide* demethylase

activity, by removal of the methyl moiety from 5mdC, was initially assigned to MBD2 (Battacharya et al., 1999). However, evidence for MBD2 5mdC glycosylase activity could not be independently reproduced (Wade et al., 1999)(Ng et al., 1999), raising technical concerns over the original experimental approach. Additional questions revolve around the feasibility of this particular reaction. Although cleavage of carbon-carbon bonds is known to be catalysed by enzymes in the thymidine salvage pathway and cholesterol synthesis pathway, this process requires enormous catalytic power (Smiley et al., 2005)(Lepesheva et al., 2007). MBD2 is known to stably bind methylated cytosines (Ng et al., 1999)(Hendrich & Bird, 1998) and thus, it is unclear how this affinity and efficient catalysis can be resolved. Serious concerns over the biological relevance of MBD2-mediated DNA demethylation are further highlighted by the analysis of *Mbd2* knockouts; *Mbd2*^{-/-} mice are viable with no obvious methylation (or demethylation) defects (Hendrich et al., 2001)(Santos et al., 2002).

Meanwhile, 5mdC glycosylase activity has been attributed to MBD4. Sequence homology between MBD4s N- and C-terminal domains and bacterial glycosylase and endonuclease domains (including the *E. coli* MutY and *Methanobacterium thermoautotrophicum* MIG) suggest that mechanistically MBD4-catalysed DNA demethylation resembles the direct removal of 5mdC by the glycosylase lyase activity of DME and ROS1. However, further assays detected neither glycosylase nor endonuclease activity in recombinant MBD4 and MBD4 is not necessary for the genome-wide active DNA demethylation of the paternal genome (Santos et al., 2004). Therefore, although MBD4-mediated active DNA demethylation of specific loci is a continuing area of research, it is widely accepted that these mechanism do not involve the thermodynamically improbable direct excision of the methyl group.

An early attempt to isolate a putative DNA demethylase identified RNA-dependent 5mdC-glycosylase activity in the lysates of chick embryos (Jost, 1993). The protein component of this factor was later identified as the base-excision repair protein thymidine DNA glycosylase (TDG). TDG specifically catalyses the removal of thymidine or uracil from G:T and G:U mismatches respectively (Neddermann et al., 1994). 5mdC glycosylase activity, however, is 30-40 fold lower than its classical thymidine glycosylase activity, and some have suggested that even this modest enzymatic activity is due to impurities in the substrate DNA. Nonetheless, TDG does

appear to play a role in active DNA demethylation. Given its high enzymatic activity against T:G mismatches, it is possible that TDG-mediated excision of the base acts downstream of 5mdC deamination.

1.5.6 DNA demethylation by deamination and DNA repair

Somewhat surprisingly, the *de novo* DNA methyltransferases DNMT3A and DNMT3B were early candidates 5mdC deaminases. The mechanism of DNA methylation has been extensively studied in bacterial DNA methyltransferases (MTases), including M.HpaII and M.SssI, and is believed to be highly conserved in their mammalian orthologues (Fig. 1.3). Methylation involves the eversion, or base-flipping, of the target cytosine from the DNA helix and its insertion into the DNA methyltransferase catalytic cleft. Nucleophilic attack of carbon-6 by the thiol group of the cysteine in the strictly conserved proline-cysteine-glutamine (PCQ) motif covalently anchors the enzyme to its target base. Simultaneous protonation of nitrogen-3 by the glutamate of the ENV motif produces an activated 4,5 enamine intermediate. Electron displacement into the pyrimidine ring activates the otherwise inert carbon-5 towards electrophilic attack on the sulfonium-linked methyl group of the universal methyl-donor (SAM). Elimination of the C5 proton and resolution of the covalent intermediate produces 5-methylcytosine and the homocysteine SAH. However, in the absence of SAM, the 4,5 enamine intermediate is proposed to destabilise the C4 endocyclic amine, resulting in accelerated oxidative deamination of the 5mdC base. Evidence for the involvement of the *de novo* Dnmts in active DNA deamination include the correlation between Dnmt-promoter occupancy and DNA demethylation during the cyclical methylation state changes of the oestradiol-oestrogen receptor target gene pS2 (Metivier et al., 2008). Additionally, *de novo* DNMTs are known to interact with DNA glycosylases, therefore physically linking deamination and T:G glycosylase activity. The physiological relevance of this reaction is yet to be proven.

A

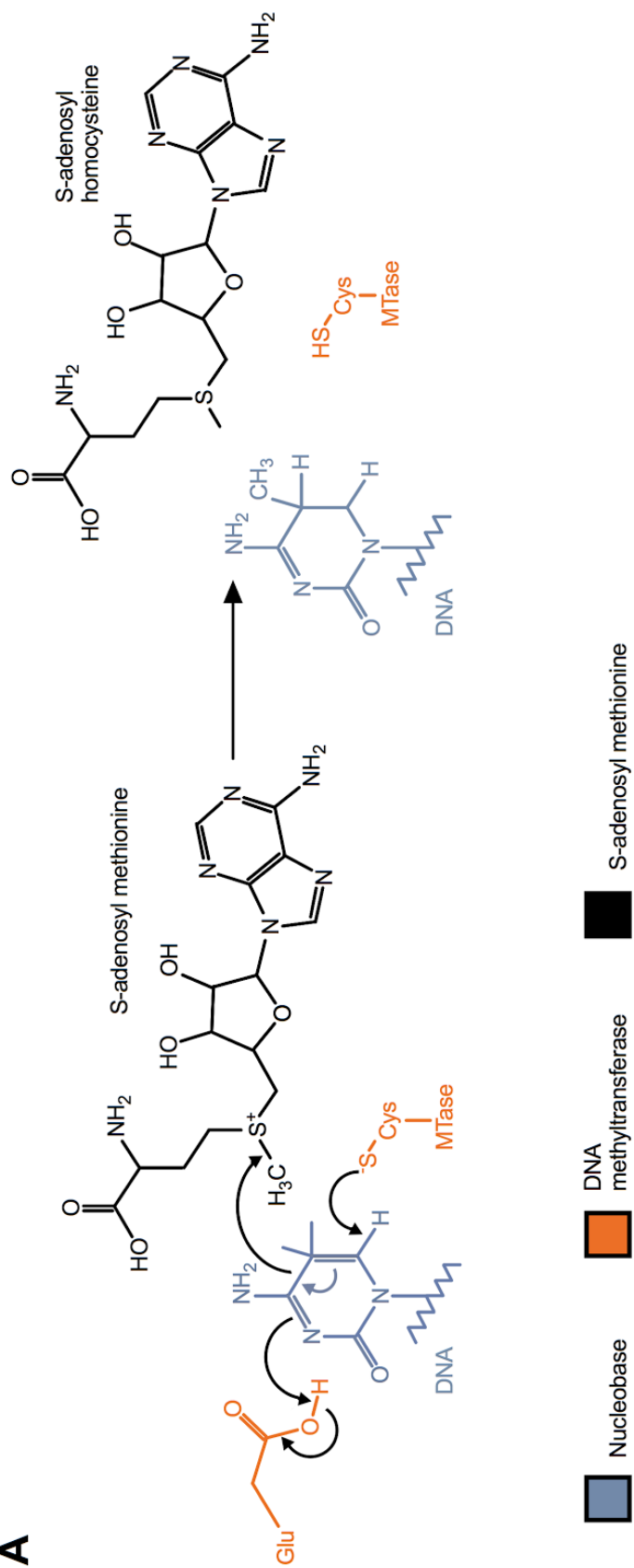


Figure 1.3 Mechanism of DNA methylation by MTase and SAM

Activation induced cytidine deaminase (AID) is perhaps best known for its role in immunoglobulin diversification. Shown to act as a cytosine deaminase *in vitro* (Bransteitler et al., 2003), in *E.coli* (Petersen-Mahrt et al., 2002), yeast (Poltoratsky et al., 2004) and in vertebrate cells (Honjo et al., 2002)(Martin et al., 2002), diversification is widely believed to be a result of cytosine deamination and repair of the resulting G:U mismatches. The substrate specificity of AID, and the related protein APOBEC1, was extended by experiments demonstrating deaminase activity towards 5mdC *in vitro* and in *E.coli* (Morgan et al., 2004). Initially both AID and APOBEC1 were reported to be highly expressed during periods of global epigenetic reprogramming (i.e. in PGCs and zygotes)(Morgan et al., 2004), suggestive of a role in active DNA demethylation via deamination and DNA repair (Morgan et al., 2004). Notably however, deamination of 5mdC is thought to be 10-fold less efficient than deamination of unmodified cytosine.

A large body of evidence supporting the role of the AID/APOBEC family as DNA demethylases stems from studies of zebrafish embryonic development. Initial experiments demonstrated that exogenously methylated DNA (M-DNA) injected into zebrafish oocytes would demethylate in a replication independent (and therefore active) manner, between 1 and 12 hours-post-fertilisation (hpf) (Collas, 1998). A decade later, injection of M-DNA was shown to elicit the concurrent demethylation of genomic DNA (Rai, 2008), and upregulation of *Aid*, *Apobec2a* and *Apobec2b* mRNA expression (Rai, 2009). Overexpression of AID (or the thymidine glycosylase MBD4) also resulted in the global loss of DNA methylation whilst knockdown of AID introduced aberrant methylation at CpG islands; suggesting that deamination of 5mdC to thymidine by AID, followed by MBD4-mediated base-excision repair were key steps in active DNA demethylation during early zebrafish development.

Like many pioneering studies in the field of active DNA demethylation, these results have been fraught with controversy, particularly in regards to experimental reproducibility. In different hands, injection of exogenous M-DNA into zebrafish oocytes resulted in neither global or locus specific DNA demethylation (measured by methylation sensitive HpaII digestion and bisulphite sequencing). Therefore, the role of AID and other cytidine deaminases in active DNA demethylation remains to be clarified. More recent models suggest that processive

long patch BER and MMR of deaminated cytosine and 5-methylcytosine (Grin et al., 2016)(Shibutani et al., 2014) involving the glycosylase UNG2 could simultaneously demethylate multiple methyl-CpGs (Santos et al., 2013)(Franchini et al., 2014)(Wu et al., 2014)(Pfaffeneder et al., 2014)(Xue et al., 2016).

1.5.7 DNA demethylation by oxidation and DNA repair

In *Trypanosoma brucei* (*T. brucei*), glycosylated 5-hydroxymethyluracil (also known as 5hmU or Base J) regulates surface glycoprotein expression in order to evade the immune response. Biosynthesis of Base J depends on the oxidation of thymidine to 5-hydroxymethyluracil by two Fe(II)- and alpha ketoglutarate-dependent enzymes (JBP1 and JBP2). The initial oxidation product, 5hmU, can be further oxidised to 5-formyluracil (5fU) and 5-carboxyluracil (5caU). Bioinformatic analysis identified the Ten-Eleven Translocation (TET) family as mammalian orthologs of JBP1 and JBP2 (Tahiliani et al., 2009). These large multidomain enzymes have since emerged as the most plausible candidate implicated in active DNA demethylation in mammals.

There are 3 members of the TET family (TET1-3). Overexpression of TET1 results in significant genome-wide loss of 5mdC, suggesting that unlike JBP1 and JBP2 in the presence of its two cofactors, TET1 primarily targets modified cytosine. In the presence of their two cofactors (Fe(II) and 2-oxoglutarate) TET proteins were also discovered to catalyse the successive oxidation of 5mdC to 5hmC, 5fC and 5CaC (Fig. 1.4). This reaction preferentially occur within a CpG context.

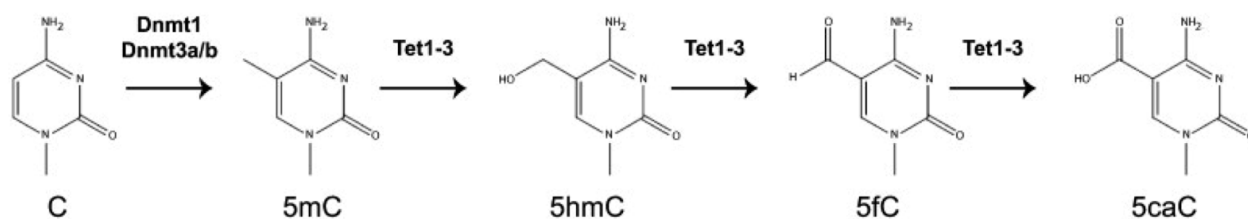


Figure 1.4 Overview of Tet-catalysed 5mdC oxidation
From (Hill, Amouroux & Hajkova, 2014)

Following oxidation, the 5mdC derivatives may be lost by either passive dilution or direct excision of 5fC and 5CaC. Thermodynamically, the activation barrier for the removal of an oxidised derivative, for example by decarboxylation (Schiesser et al., 2012), is substantially lower and therefore is more feasible than the removal of an unmodified methyl group. However, enzymes able to catalyse this reaction await discovery. Additionally, *in vitro* abundant SAH or absence of SAM can promote the removal of oxidised 5mdC derivatives by DNA methyltransferases. However, the biological relevance of the reverse DNMT reaction is unknown given that SAM is a key cellular metabolite and it is unclear whether the global or local absence of SAM is physiologically possible.

Removal of the oxidised base by base excision DNA repair (BER) is a more plausible explanation and is far less controversial. Since TDG primarily excises T from T:G mismatches, initial studies assumed that oxidised 5mdC derivatives bases would first be deaminated to T. To this end, AID/APOBEC were considered as catalysts for deamination. However, *in vitro* and *ex vivo* evidence for deamination of 5hmC by Aid/Apobec does not exist and *in vivo* studies showed no significant developmental defects associated with AID/APOBEC deficiency.

TDG alone has been shown to specifically recognise and excise 5fC and 5CaC (but not 5mdC or 5hmC) *in vitro* and *in vivo* (He et al, 2011)(Maiti & Drohat, 2011)(Zhang et al., 2012). Knockout of *Tdg* is embryonic lethal due to aberrant promoter methylation and decreased expression of the HOX family transcription factors;

highlighting the importance of TDG for the regulation of DNA methylation levels, gene expression and ultimately embryonic development. Additionally, knockdown of TDG leads to elevated 5CaC in ESCs and the parallel overexpression of both TET and TDG in HEK293 cells leads to genomic depletion of 5fC and 5CaC. Increased TDG glycosylase activity towards 5fC and 5CaC compared to 5mdC is most probably due to the altered steric and electronic properties causing subtle structural distortions of the oxidised bases, promoting the destabilisation of the N-glycosidic bond (Williams & Wang, 2012)(Raiber et al., 2015).

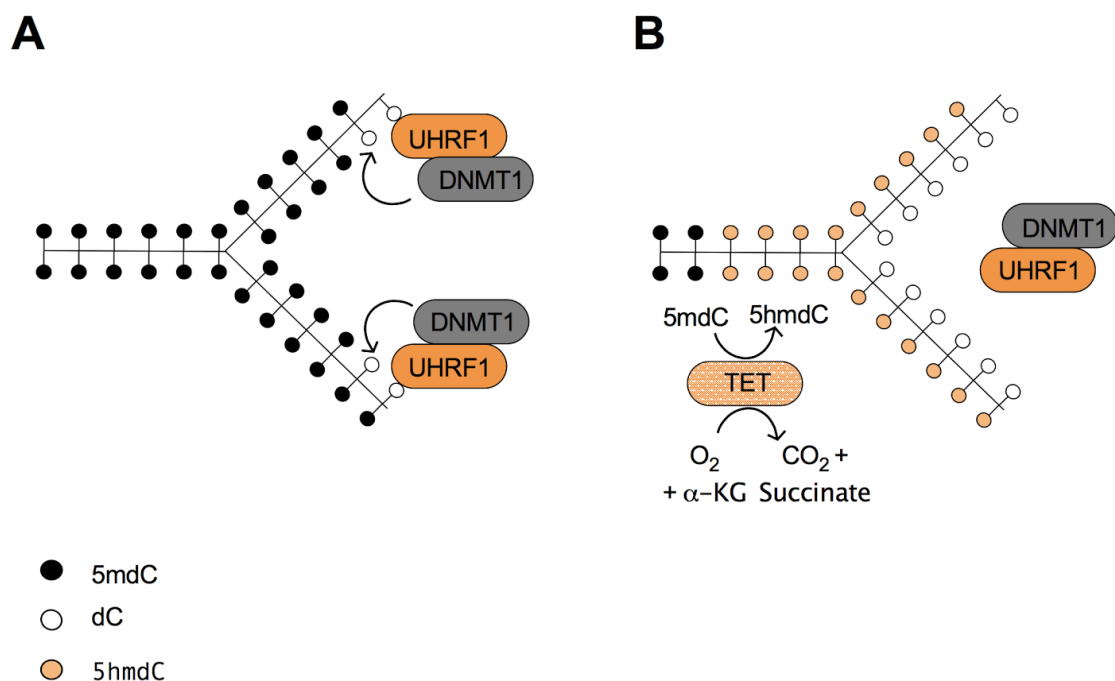


Figure 1.5 5-hydroxymethylcytosine and passive DNA demethylation

(a) DNMT1 ensures that DNA methylation states are faithfully maintained following DNA replication

(b) Oxidation of 5mdC by TET proteins produces 5hmdC. Although UHRF1 interacts with 5hmdC, DNMT1 has low affinity for oxidised 5mdC derivatives. Therefore, TET-triggered 5mdC oxidation may lead to passive replication-dependent DNA demethylation.

Tet proteins have consistently been found to colocalise at specific genomic loci with a host of other BER-proteins; including MBD4, PARP1, XRCC1 and NEIL1-3). It is likely that these proteins act downstream from TDG-mediated glycosylation to repair the abasic site. For example, bifunctional DNA glycosylases have

been shown to interact with TDG during the reactivation of silent reporter genes in differentiated cells and mESCs (Muller et al., 2014)(Schomacher et al.,2016) and *Xenopus* embryogenesis was found to depend on TET3, TDG and NEIL1 and 2 (Schomacher et al., 2016). Recent models propose that NEIL glycosylases are non-catalytic partners which facilitate the dissociation of TDG from abasic sites following the excision of 5fC and 5CaC so as to allow strand incision by AP endonuclease 1 (APE1). Whether these proteins have an enzymatic role or regulatory role is under investigation. Combined, these experiments somewhat reconcile the need for TDG with insights into TET-mediated oxidation. Importantly, oxidation coupled to base-excision repair is the first biologically and biochemically validated complete pathway for active DNA demethylation.

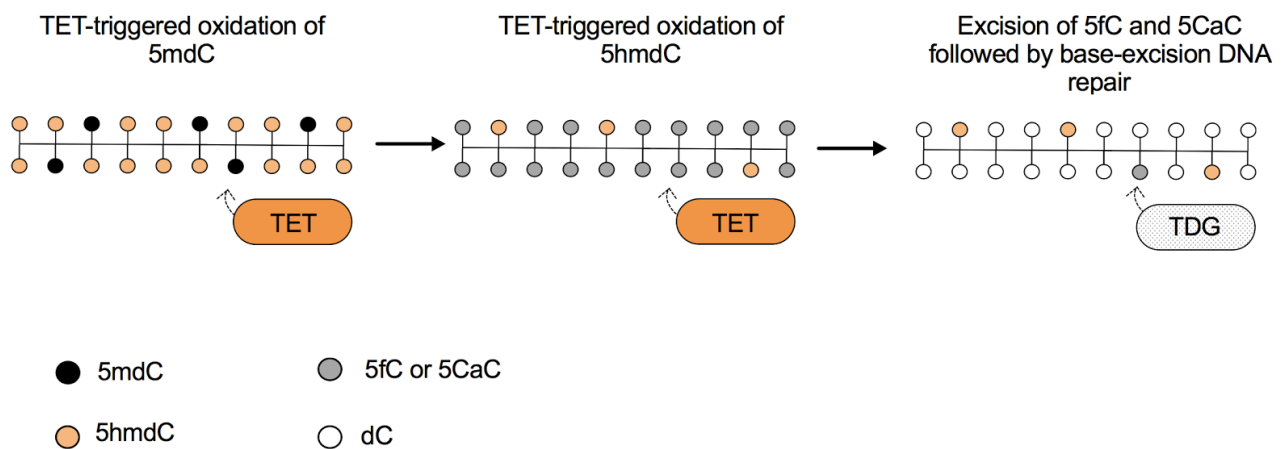


Figure 1.6 5hmC and active DNA demethylation

Iterative oxidation of 5mC produces firstly 5hmC and then 5fC and 5CaC. Both 5fC and 5CaC are substrates for the DNA glycosylase TDG. Excision of 5fC and 5CaC by TDG is proposed to activate the base-excision DNA repair pathway by generating abasic lesions. Recruitment of BER proteins, including poly (ADP-ribose) polymerase, DNA polymerase, DNA ligase and the scaffold protein XRCC1 is required to repair the abasic site by the addition of an unmodified cytosine.

1.5.8 The supporting cast

The most plausible models for active DNA demethylation involve multiple steps. Whilst there may only be a few proteins directly involved in the modification, excision and repair of the CpG site, many other proteins, including the Growth arrest and DNA damage protein family (Gadd45) are likely to facilitate or increase the efficiency of DNA demethylation without being the catalyst themselves.

GADD45A and GADD45B have repeatedly been associated with DNA demethylation; coimmunoprecipitation show a direct physical interaction between GADD45A, AICDA and TDG in cultured human cells (Barretto et al., 2007) whilst a similar tripartite consisting of AICDA/APOBEC, MBD4 and GADD45A exists in zebrafish (Rai et al., 2008). Expression patterns of GADD45 proteins also forge an interesting link between DNA methylation and psychiatric disorders. Compared to non-psychiatric controls, GADD45B and GADD45G (along with MBD glycosylase, TET1 and APE1) are elevated in patients with psychotic illnesses. More specifically, GADD45g is elevated in the prefrontal cortex of patients with schizophrenia whilst GADD45b expression is increased in the parietal cortices of patients with psychosis. Interestingly, increased expression of Dnmts is also observed in patients with psychosis, perhaps explaining the need for increased demethylase activity. Alternatively, the elevated expression of BER DNA repair proteins may simply be a response to greater genotoxic stress caused by the administration of psychoactive drugs. The current model for GADD45 involvement in active DNA demethylation proposes that the protein binds to and focuses the activity of glycosylases and deaminases to specific gene promoters.

1.6 Dynamic DNA methylation during development

1.6.1 DNA demethylation during preimplantation development

The dynamic nature of DNA methylation is most notable during the periods of global epigenetic reprogramming that occur in early preimplantation development and during the differentiation of primordial germ cells (see Chapter 8). Following the decondensation of the paternal pronucleus, the paternal genome undergoes global DNA methylation erasure. Accumulation of 5hmC is also observed on the paternal genome, coinciding with the binding of TET3, the only TET protein expressed in zygotes, and the activation of the BER DNA damage pathway (Wossidlo et al., 2011)(Zhang, 2012)(Gu, 2011)(Iqbal et al., 2011). Abolishing TET3 activity by knockdown or conditional knockout has been reported to impact the dynamic 5mdC loss and 5hmC appearance (Inoue et al., 2011)(Iqbal et al., 2011)(Wossidlo et al., 2011)(Gu et al., 2011); apparently mechanistically linking DNA demethylation and TET-mediated 5mdC oxidation. However, more recent studies

by our group have conclusively proven this to be incorrect (Amouroux et al., 2016). Immunofluorescent staining has shown that 5mdC is lost by the early pronuclear (PN) stage 3. Direct coupling between DNA demethylation and hydroxylation should result in a concomitant increase in 5hmC. However, oxidised 5mdC derivatives (5hmC, 5fC and 5CaC), do not accumulate until PN4. These processes can be forcibly uncoupled using a small molecule inhibitors of 2OG-dependent oxygenases (dimethyloxallyl, DMOG) and the Fe²⁺ chelator deferoxamin (DFX), resulting in the complete loss of 5hmC signal *without* affecting the prior loss of 5mdC. 5hmC is also significantly decreased when the zygotes are exposed to the DNMT inhibitor 5-Azacytidine; indicating that the formation of 5hmC is somewhat dependent on newly deposited *de novo* methylation. Furthermore, 5hmC is reduced in zygotes when both *Dnmt3a* and *Dnmt1* are deleted from the oocyte, suggesting that this *de novo* activity is maternally derived. Mitotic chromosome spreads of blastomeres from the 1-, 2- and 4-cell stage suggest that, following oxidation of *de novo* 5mdC, 5fC and 5CaC is lost by passive, replication-dependent dilution (Inoue et al., 2011).

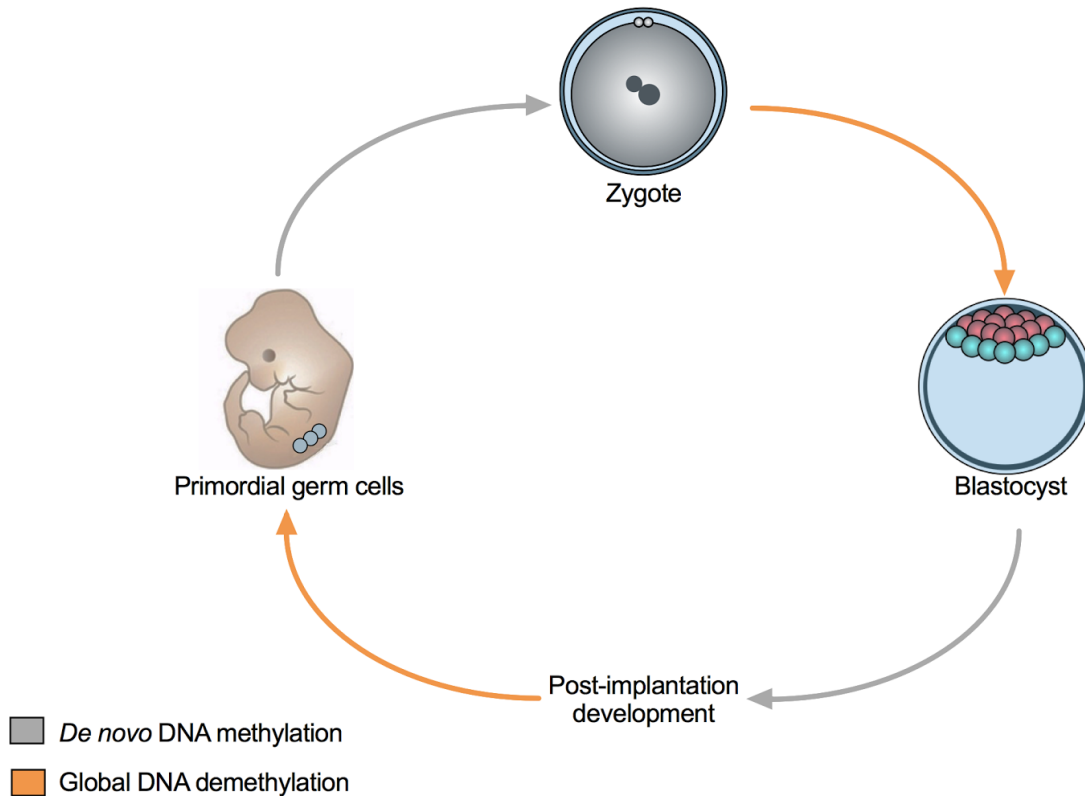


Figure 1.7 DNA methylation dynamics during embryonic development

Following fertilisation, both maternal and paternal genomes undergo global DNA demethylation. The timing of DNA demethylation suggests that in each instance different mechanisms are responsible for the loss of 5mdC; active for the paternal genome and passive for the maternal genome. Post-implantation development is characterised by a global wave of *de novo* DNA methylation. This period is associated with lineage commitment and differentiation into somatic cells. Prior to gastrulation, cells of the proximal epiblast are induced to begin germ cell development. This entails the genome-wide loss of 5mdC. Following primordial germ cell development, sex-specific imprints and DNA methylation patterns are re-established.

Therefore, TET-triggered hydroxylation of 5mdC may act as a mechanism for removing aberrant sporadic 5hmdC on an otherwise hypomethylated genome. This model can explain why, despite very low levels of TDG expression in zygotes (Hajkova et al., 2010)(Tang et al., 2011), DNA demethylation is possible. Clearly, the mechanism behind the active DNA demethylation of the paternal genome remains elusive. Whilst DNA demethylation, 5mdC hydroxylation, and the BER DNA repair pathway all appear to be involved, their relationship is complex and perhaps indirect.

1.6.2 DNA demethylation in pluripotent stem cells

DNA methylation has been studied extensively in mouse embryonic stem cells (mESCs). Although mESCs retain pluripotency in the complete absence of 5mdC (Tsumura et al., 2006), dynamic methylation is required to silence the pluripotency transcriptional circuit upon differentiation (Jackson et al., 2004). In wild type mESCs, global abundance of 5mdC is dependent on the culture condition (Leitch et al., 2013)(Ficz et al., 2013)(Habibi et al., 2013). Conventional FCS-containing culture conditions are associated with DNA hypermethylation whilst mESCs cultured in FCS-free medium in the presence of two ERK and GSK3 β inhibitors (known as 2i medium) are globally hypomethylated.

TET proteins are proposed to control the methylation state of poised and active gene promoters and developmentally important enhancers (Hon et al., 2014)(Lu et al., 2014)(Serandour et al., 2012) (Koh et al., 2011)(Wu et al., 2011)(Huang et al., 2014)(Song et al., 2012)(Pastor et al., 2011)(Sun et al., 2015); maintaining the expression of pluripotency genes and stabilising the poised state by protecting them from *de novo* methylation and long term silencing. These observations are consistent with the pivotal role of TET1 and TET2 in iPSC reprogramming; *Tet1* mRNA overexpression enhances reprogramming efficiency by promoting DNA demethylation and expression of *Oct4*, and can even replace *Oct4* in the OSKM reprogramming cocktail. Meanwhile, TET2 is recruited to the promoters of the pluripotency-related genes *Nanog* and *Esrrb* following OSKM transduction, activating their transcription. Conflicting reports show that knockdown of either *Tet1* or *Tet2* mRNA abrogate mESC maintenance and skew differentiation in favour of the trophectoderm and primitive endoderm lineages. It is likely that these opposing observations reflect differences in the genetic background or culture conditions. Whilst genetic ablation of *Tet1* and *Tet2* is compatible with embryonic development, *Tet(1-3)* triple knockout embryos are unable to support differentiation (Dawlaty, 2014) and it is widely accepted that the knockout of all three TET proteins cannot support embryonic development.

In agreement with previous findings, abrogation of TDG-activity in mESCs causes significant enrichment of 5fC and 5CaC at distal gene regulatory elements (Song et al., 2013)(Shen et al., 2013)(Neri et al., 2015). These regions are commonly co-occupied by TET1 and TDG in wild type cells (Neri et al., 2015) and the direct

interaction between these two proteins and their accessory factor (GADD45A) have repeatedly been shown (Muller et al., 2014)(Weber et al., 2016)(Keinhofer et al., 2015)(Li, 2015), suggesting that the hypomethylation of developmental enhancers is dependent on Tet-mediated oxidation and TDG-mediated base excision. Currently there is little evidence for substantial involvement of additional pathways of active DNA demethylation in pluripotent and differentiating cells; AID deficiency in mice has no overt effect on embryonic development or the establishment of cell type specific DNA methylation patterns (Fritz et al., 2013), and knockout of DNA glycosylases other than TDG in mice do not interfere with developmental capacity.

1.6.3 DNA demethylation in differentiated cells

Despite the supposed stability of DNA methylation, maintenance and *de novo* Dnmts are expressed in virtually all somatic, terminally differentiated cells types; suggesting that maintenance of steady state 5mdC requires ongoing DNA methylation. Transcription is likely to play a pivotal role in reshaping DNA methylation (Morano, 2014), evidence of which can be seen in the cyclical fluctuations of gene promoter methylation upon transcriptional stimulation by nuclear and hormone receptors (Kangaspeska et al., 2008)(Metivier et al., 2008). In these cells, gene activation is accompanied by an overall loss of DNA methylation and the binding of DNMTs, GADD45A, TET and BER proteins including TDG and APE1. Additionally, binding of CTCF and REST to CpG poor distal regulatory elements coincides with their locus-specific DNA demethylation. Given the local enrichment of TET hydroxylases and oxidised 5mdC derivatives, this local DNA demethylation process is thought to involve TET-mediated oxidation and BER DNA repair. Inactivation of all three TET enzymes in differentiated human cells causes DNA hypermethylation (Grosser, 2015), further supporting the idea that operational TET-based active DNA demethylation pathways are, to some extent, important in all cell types. In agreement with our group's model of TET-mediated protection of hypomethylated DNA from aberrant *de novo* methylation, TET1 is known to prevent the spreading of DNA methylation into CGIs in human HEK293 cells (Jin et al., 2014) and oxidise 5mdC at sites of DNA repair (Kafer et al., 2016)(Morano et al., 2014).

1.7 Pluripotency

1.7.1 Defining the naive state

Pluripotent cells are defined by two fundamental properties: indefinite, rapid symmetrical cell divisions and the ability to differentiate into specialised cells of the three germ layers (the ectoderm, mesoderm, and endoderm) and the germline. Multiple assays assess functional pluripotency as a measure of developmental potential: *in vitro* differentiation following the addition of differentiation inducing cytokines or small molecules, the capacity to form teratocarcinomas following subcutaneous injection, the high tissue contribution and germline transmission in chimaeras, and finally tetraploid (4n) complementation. The latter two assays are considered the definitive tests for pluripotency; in tetraploid complementation, 4n donor blastocysts are unable to develop past mid-gestation and thus embryonic development relies on the widespread contribution of *ex vivo* cultured pluripotent cells, and similarly the injection of a pluripotent cell into the blastocyst provides proof that a single cell is able to reenter embryonic development and generate multiple different cell types. Crucially, this developmental capacity must be restrained in mice to prevent tumorigenesis. *In vivo*, pluripotency emerges during the development of the totipotent zygote into a blastocyst.

1.7.2 From cleavage to gastrulation

Mammalian embryonic development has many quirks, foremost being the reliance on extraembryonic tissue for implantation, nutrient transfer and embryonic patterning. The totipotent zygote is the only cell capable of a) producing all extra- and embryonic tissues and b) coordinating development to produce viable offspring. Even so, zygote development is highly restricted; dissociated blastomeres differentiate into either microblastocysts of ICM-like cells or trophoblast vesicles (Tarkowski and Wroblewska, 1967, Ziomek and Johnson, 1980). Thus, instead of representing just a gradual decline in developmental potency, pre-gastrula development is a crucial preparatory stage in the acquisition of the true ground state pluripotency present in the epiblast.

Pre-gastrula development essentially consists of two cell fate decisions; each based on specific signalling pathways and solidified by mutually reinforcing or antagonising transcriptional programmes. As shown by transplantation studies, prior to the 8 cell stage all cells are identical in relation to their polarity, position within the conceptus, and cell-cell contacts. As such, they also have identical developmental capacity and can produce either embryonic or extraembryonic tissue (Gardner, 1998)(Hillman et al., 1972). At the 16-cell stage, morula compaction and the accumulation of conserved cell fate determinants (polarity proteins) precedes cell diversification. Differentiative cell divisions based on cell polarity form the first progenitors of the trophectoderm and the ICM (Fig. 1.8). Whilst functional HIPPO signalling and SOX2 expression is specifically maintained in the basal cells (Wicklow et al., 2014), repression in the outer cells facilitates the assembly of the YAP/WWTR1/TEAD4 complex. This complex promotes the transcriptional activation of the trophectoderm genes *Cdx2* and *Gata3* and the silencing of *Sox2*. Expression of *Cdx2* precedes the silencing of the pluripotency master regulators *Oct4* and *Nanog* and the upregulation trophectoderm-specific genes including *Eomes* (Carey et al., 2014)(Strumpf et al., 2005)(Chen et al., 2009). Thus, the first cell fate decision delineates the polarised epithelium of the trophectoderm and the apolar inner cell mass (ICM).

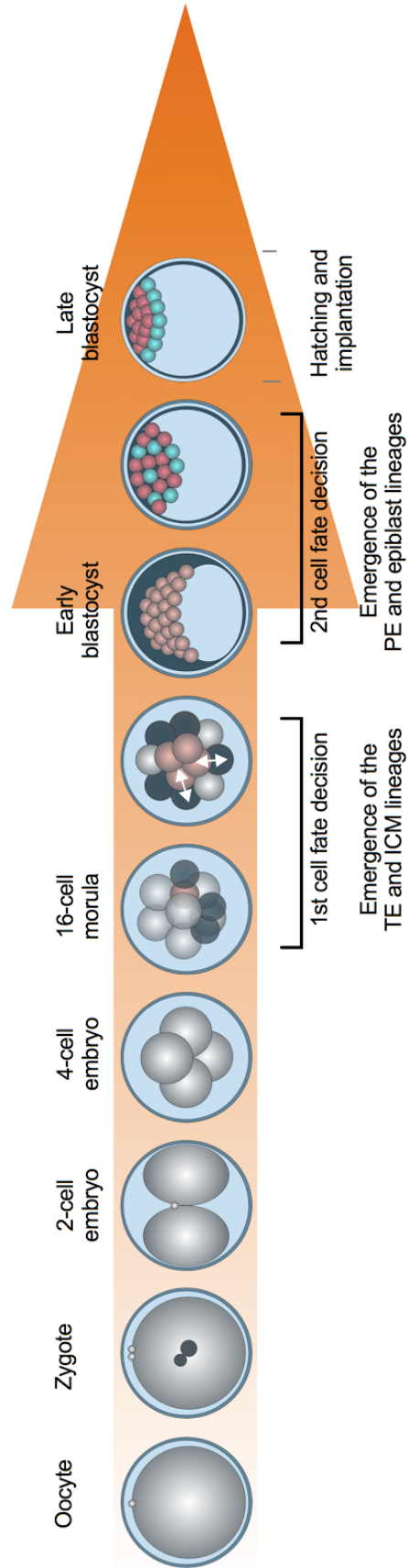


Figure 1.8 Schematic of preimplantation development and the regulation of cell fate decisions

The second cell fate decision, defining differentiation into either epiblast or primitive endoderm (PE) was initially believed to be governed by a given cell's position within the ICM (Gardner, 1983). The mosaic expression of GATA6 and NANOG within the ICM, however, suggests otherwise (Chazaud et al., 2006)(Rossant et al., 2003). Instead, FGF signalling regulates PE formation. *Fgf4* is expressed in epiblast precursor cells whilst the receptor *Fgfr2* is downregulated (Rappolee et al., 1994)(Frankenberg et al., 2011)(Guo et al., 2011). Meanwhile, PE precursor cells maintain *Fgf2r* gene expression but downregulate the signalling ligand *Fgf4* (Guo et al., 2010). Therefore, the FGF4/MAPK intracellular signalling pathway is specifically activated in PE precursor cells. FGF4/MAPK signalling initiates the transcriptional expression of primitive endoderm genes and simultaneously represses the expression of epiblast genes, ultimately separating the primitive endoderm and epiblast cell fate by positive feedback and mutually exclusive transcriptional programmes (Kang et al., 2013)(Chazaud et al., 2006). For instance, the expression of the pluripotency related gene *Nanog* and the PE-marker *Gata6* are mutually exclusive. Abolishing *Nanog* expression induces *Gata6* gene expression throughout the entire ICM (Frankenberg et al., 2011) whilst enforced *Gata6* expression in mESCs results in the down-regulation of pluripotency related genes, including *Nanog* (Fujikura et al., 2002)(Shimosato et al., 2007). Primitive endoderm cells are highly motile due to the expression of laminin and DAB2 and therefore are able to migrate towards the blastocoelic cavity (Morris et al., 2002). Collectively this preferred mechanism for the second cell-fate decision is known as the *cell-sorting* model.

1.7.2 Capturing pluripotency - Teratocarcinomas and EC cells

The high incidence of spontaneous testicular teratocarcinoma in inbred mice provided the first proof that a self-renewing cell population existed *in vivo* (Stevens & Little et al., 1954). In 1967, two seminal papers explored the unique developmental origin and potency of teratocarcinomas (Stevens, 1967)(Pierce, 1967). Concluding teratocarcinomas are naturally derived from primordial germ cells, it was shown that these benign tumours can be experimentally induced by the ectopic transplantation of genital ridges. A single teratocarcinoma cell is able to produce a tumour consisting of multiple different cell types, and is therefore classified as multipotent (Mintz & Illmensee, 1975).

Ex vivo, these cells can be maintained as embryo carcinoma (EC) cells. However, serial passaging of EC cells is known to hamper their developmental potency. This can be assessed following subcutaneous injection, where the resultant tumours characteristically contain a high proportion of poorly differentiated cells. EC cells can also be induced to differentiate into cystic embryoid bodies with a complex mixture of embryonic cell types. The similarities between the structure of the embryoid body and the blastocyst, the capacity to contribute to chimaeras, and the panel of shared cell surface markers suggests that similar pluripotent cells must also exist in the blastocyst (Artzt et al., 1973)(Jacob, 1978).

1.7.3 Capturing pluripotency - Embryonic stem cells

It therefore seemed plausible that pluripotent cells could be isolated directly from the embryo, without the intervening tumour stage. Although the initial attempts to derive these cell lines were wholly unsuccessful (Sherman, 1975)(Martin, 1975), the optimisation of culture conditions to support the growth of clonal EC cells eventually facilitated the isolation of mESCs. Specifically, mESCs were derived and maintained in foetal calf serum (FCS) containing medium on a layer of mitotically inactivated fibroblasts (feeder cells) which provide trophic factors to support the growth and maintenance of pluripotency. Crucially, these mESCs were pluripotent; self-renewing indefinitely and generating chimaeras with significant contribution to all somatic cell types and the germline. Upon microdissection of E4.5 blastocysts and individual culture of the trophectoderm, the epiblast and the endoderm, only the epiblast cells produced colonies, providing proof that the epiblast was the origin of mESCs (Brook & Gardner, 1997).

Feeder cells inhibit differentiation by both contact dependent and independent mechanisms (Isacke et al., 1988) and the removal of the feeder layer resulted in spontaneous differentiation (Smith and Hooper, 1983)(Heath and Smith, 1988). Even so, maintaining a pluripotent state in the presence of feeder cells was challenging. Furthermore, feeder cells created a complex and undefined molecular environment which limited and confounded attempts to understand the molecular mechanisms underlying mESC self renewal and differentiation (Heath and Smith, 1988). Therefore, research groups set out to develop appropriate, defined culture conditions for the maintenance of undifferentiated mESCs. Media conditioned by feeder cells, such as buffalo rat liver (BRL) cells, had potent anti-differentiation properties. This effect was attributed to the

production of a single cytokine; Leukaemia Inhibitory Factor (LIF) (Smith & Hooper, 1987)(Zwaka & Thomson, 2005) (Williams et al., 1988)(Stewart et al., 1992). LIF is expressed by all feeder cells and co-culture of mESCs and *Lif*-null feeders does not efficiently support ESC self-renewal (Stewart et al., 1992). Today, maintenance of germline competent ESCs in feeder-free conditions with appropriate batches of FCS and supplemented with recombinant LIF is commonplace (Nichols et al., 1990). These FCS/LIF culture conditions are considered the conventional condition for ESC culture.

1.7.4 LIF/Stat3/PI3K signalling and pluripotency in embryonic stem cells

The pivotal role of LIF in ESC survival created an urgent need to explore the intracellular signalling cascades linking LIF ligand binding and pluripotency. Extracellular LIF binds to the IL-6 signalling transmembrane receptor (LIF-receptor, LIF-R); mediating its heterodimerisation with the closely related signal transducer glycoprotein 130 (gp130)(Davis et al., 1993) (Fig 1.9). Chimeric gp130 receptors which are unable to engage STAT3 do not support mESC self renewal. Furthermore, related cytokines which also act via the gp130 receptor are able to replace LIF and sustain pluripotency (Yoshida et al., 1994). Gp130 associates with the tyrosine phosphatase SP-2; a positive effector of the ERK signalling pathway. However, mutation to the phosphotyrosine docking site (tyrosine 118) was shown to prevent activation of ERK signalling without affecting the pluripotency. Therefore, LIF signalling via the ERK pathway is not responsible for the maintenance of pluripotent mESCs (Burdon et al., 1999).

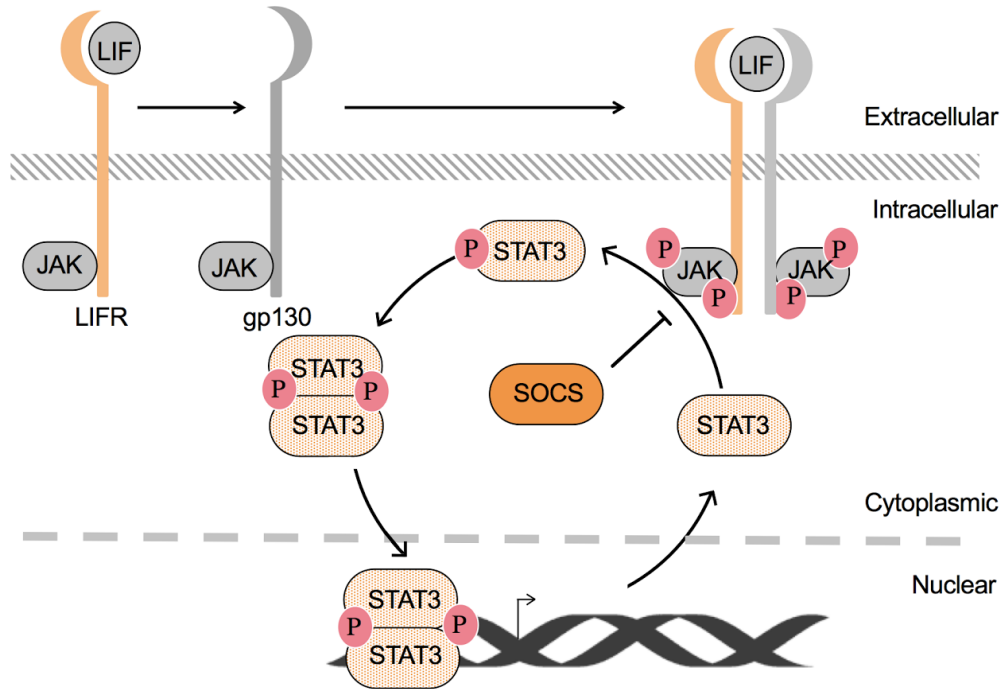


Figure 1.9 LIF induces the JAK/STAT signalling cascade

Binding of LIF to the IL-6 receptor LIFR induces heterodimerisation with gp130 and the activation of Janus-associated kinases (JAKs). JAKs phosphorylate the IL-6 receptors cytoplasmic domain, generating docking sites for STAT3. Once phosphorylated, STAT3 dimerises and translocates to the nucleus, where it regulates its target gene expression. Several target genes, including SOCS, negatively regulate the JAK/STAT signalling cascade via a negative feedback loop.

Instead, LIF primarily activates the signal transducer and activator of transcription (STAT) cascade (Fig. 1.9). The catalytically inactive membrane proximal region of IL-6 receptors is constitutively bound to non-receptor Janus-associated (JAK) tyrosine kinases. Ligand binding and multimerisation brings the JAKs associated with each receptor into close proximity (Fig. 1.9), facilitating transphosphorylation and activation. JAKs subsequently phosphorylate receptor tyrosines within the Y-x-x-Q motif, forming a docking platform for proteins, including their key substrates, the STAT family of transcription factors, which contain a Src-homology 2 (SH2) domain. In mESCs, the ubiquitously expressed STAT3 is the predominant target of LIF signalling and *in vivo* is required to maintain the pluripotent ICM (Do et al., 2013). This is evidenced by the fact that activation of Stat3 is sufficient to support LIF-independent ESC self-renewal in high density bulk culture (Matsuda et al., 1999), constitutively active STAT3 mutants increases the number of OCT4 positive colonies in OSKM transfected mouse embryonic fibroblasts whilst JAK1 inhibitors completely abolishes reprogramming

(Tang, 2012) and dominant negative STAT3 causes differentiation even in the presence of LIF (Niwa et al., 1988).

STAT3/LIF signalling results in the transcriptional activation of reprogramming factors (*Klf4*, *Klf5* and *c-Myc*) and naive pluripotency factors (*Pim1*, *Pim3*, *Pramel7*, *Rhox5*, *Gbx2*, *Zfp57*, and *Tfcp2l1*). Of these, *Tfcp2l1* is likely to be an key STAT3 target (Martello et al., 2013; Ye, 2013). JAK-STAT signalling is regulated by feed forwards and inhibitory regulation. For instance, Cytokine signalling 3 (SOCS3) is a potent JAK-STAT inhibitor is transcribed following JAK-STAT activation. Eliminating SOCS3 inhibition increases JAK-STAT3 signalling (Burdon et al., 1999). However, JAK-STAT signalling also initiates the transcription of LIF-R, gp130, JAK1 and STAT3.

In vivo, Paracrine LIF signalling is important for the maintenance of the ICM; LIF is expressed from the trophectoderm whilst gp130 is specifically expressed by the ICM (Nichols et al., 1996) However, since there are no known early embryonic phenotypes resulting from the knockout of LIF, LIFR or gp130, pre-gastrulation these factors are not absolutely essential. Rather, LIF signalling is proposed to be involved in maintaining pluripotency specifically during diapause.

1.7.5 Akt signalling and pluripotency in embryonic stem cells

LIF signalling through JAK kinases also activates the PI3K-AKT pathway (Paling et al., 2004). AKT activation is known to promote self renewal and stabilise the pluripotency network in mESCs (Sun et al., 1999)(Takahashi et al., 2003)(Pritsker et al., 2006). However the mechanism behind this observation is unknown. A major target inhibited by AKT signalling is GSK3 β ; directly via phosphorylation of Ser9, and indirectly by phosphorylation independent nuclear export. GSK3 β targets both c-MYC and NANOG for ubiquitin mediated degradation, thus inhibition promotes self-renewal and pluripotency (Welham et al., 2011).

1.7.6 ERK signalling and defined culture medium

The composition of FCS is complex and undefined; containing numerous growth promoting and inhibitory factors, which varies significantly between batches. Therefore it is nigh impossible to tease apart the signalling pathways responsible for maintaining pluripotency. Instead, an ideal culture system should be fully defined. FCS-free medium was initially developed for the directed differentiation of mESCs (Ying et al., 2003), but has since been adapted for the long term maintenance of pluripotent cells. This shift towards FCS-free media is undoubtedly among the most important changes in stem cell culture in recent years.

In FCS-free culture, LIF alone cannot restrain mESCs from differentiating. Bone morphogenic proteins (BMPs), however, in combination with LIF *can* maintain mESC pluripotency (Ying et al, 2003) via the activation of Inhibitor of Differentiation (ID) genes. Given BMPs role in differentiation and patterning of the vertebrate embryo (Johansson and Wiles, 1995)(Wiles and Johansson, 1999)(Adelman et al., 2002)(Kawasaki et al., 2000), this effect is possibly counterintuitive. However, various embryonic models have shown that BMPs consistently inhibit neural differentiation by blocking *Sox1* activation. Therefore, in combination with LIF, BMPs and their effectors (SMADS) are able to replace FCS (Shi and Massague, 2003).

LIF and BMP signals negatively regulate FGF/p-ERK signalling; preventing lineage commitment. FGF4 is highly expressed in the early embryo (Ying et al., 2003). Transcriptional activation of the *Fgf4* gene is directly connected to the core pluripotency circuit via OCT4 and SOX2. Therefore, in the expanded blastocyst, *Fgf4* expression is restricted to the ICM (Yuan et al., 1995). FGF4 is also produced by mESCs. However, its expression is dispensable for the maintenance of the undifferentiated state. In fact, FGF/p-ERK signalling is known to sensitise ESCs towards instructive differentiation signals; both *Fgf4* *-/-* and *Erk2* *-/-* mESCs are severely compromised in differentiation towards neural and mesodermal lineages (Kunath et al., 2007) (Stravidis et al., 2007). In part, this may be due to the ERK-mediated phosphorylation, and subsequent proteasomal degradation, of KLF2 destabilising the naive pluripotency transcriptional network (Yeo et al., 2014). Therefore, it was suggested that small molecule inhibitors of FGF signalling and the ERK cascade could promote long-term, robust ESC self renewal. Small molecule inhibitors of either the FGF tyrosine kinases

(SU5402) or the ERK signalling cascade (PD184352 or the more potent PD0325901), in combination with LIF, can indeed relieve the reliance on FCS or BMP4. However culture with either inhibitor *without* LIF reduces the proliferative ability and ultimately is unable to maintain mESC self renewal.

Activation of WNT signalling enhances mESC propagation (Ogawa et al., 2006) (Sato et al., 2004), the ability of ESCs to reprogram somatic cells by cell fusion (Lluis et al., 2010), and increases the efficiency of reprogramming using *Oct4/Sox2/Klf4* transfected somatic cells (Stadtfield & Hochedlinger, 2010)(Marson et al., 2008). This may be unexpected given that, in the early embryo, WNT signalling via β -catenin is principally involved in defining the anterioposterior axis and specifying the mesodermal lineage during gastrulation. These apparent context and time-dependent effects are thought to depend on the dynamic expression of WNT ligands, receptors, antagonists and other signalling components.

Stabilisation of the WNT effector, β -catenin, effectively mimics canonical WNT signalling and inhibits the progression from naive to primed pluripotency (ten Berge et al., 2011)(Yi et al., 2011). The stability of β -catenin is regulated by a destruction complex consisting of the protein kinase Glycogen synthase kinase 3 (Gsk3), AXIN, APC, and CDK1 (Stamos and Weis, 2013). GSK3 phosphorylates β -catenin, marking it for ubiquitin mediated degradation and thus preventing transcriptional regulation through interaction with the TCF/LEF1 family transcription factors.

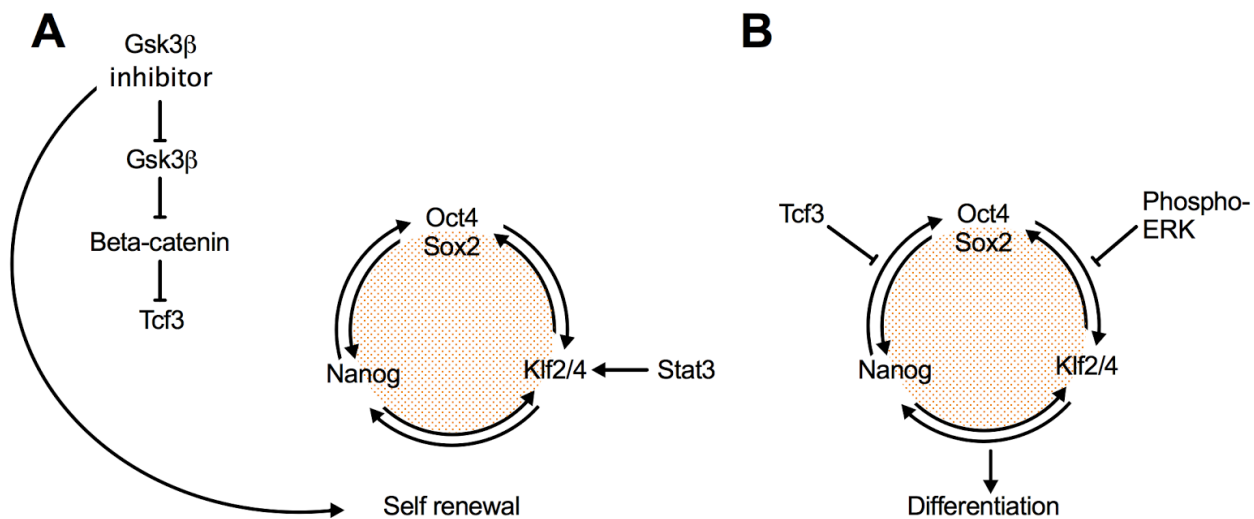


Figure 1.10 Modulation of self-renewal and differentiation by Gsk3 β and ERK inhibition.

(a) Phosphorylation of beta-catenin by Gsk3 results in proteosome-mediated degradation. Inhibitors of Gsk3 β therefore prevent beta-catenin degradation. Tcf3 is a transcriptional repressor. Interaction between Tcf3 and beta-catenin abolishes its repressive effect on multiple pluripotency-related genes. Additionally, Gsk3 β inhibition sustains mESC self-renewal via stimulatory effects on metabolic and biosynthetic processes. Additionally, Stat3 positively regulates the pluripotency circuit, primarily through Klf4. (b) In serum-free culture, absence of Gsk3 β and phospho-ERK inhibitors disrupts the pluripotency network, leading to mESC differentiation. Adapted from (Wray et al., 2010)

Initially, FCS-free medium was supplemented with the small molecule glycogen-synthase 3 (GSK3 β) inhibitor 6-bromo-indirubin-3'-oxime (BIO)(Meijer et al., 2003)(Sato et al., 2004). However, the viability of these cells was severely compromised, possibly due to cross-reactivity with cyclin-dependent kinases (Zhen et al., 2007)(Murray et al., 2004)(Bain et al., 2007). The more selective GSK3 β inhibitor, CHIR99021, enhances cell survival whilst also inducing non-neural differentiation (Ying et al., 2008). In combination with SU5402 & PD184352 or PD0325901, the differentiation protectant effect is dominant, resulting in highly proliferative, viable undifferentiated cells, even at low cell density. Therefore, inhibition of both ERK and GSK3 β maintains pluripotent mESCs in FCS-free conditions without overt negative effects by shielding the cells from differentiation-inducing signals whilst simultaneously promoting the expression of pluripotency genes. Collectively, this culture medium is known as 2i (Ying et al., 2008).

The precise mechanism by which GSK3 β inhibition and β -catenin stabilisation enhances pluripotency is also unclear. On the one hand, β -catenin may act as a transcriptional activation in complex with OCT4 and the weakly expressed TCF1 (Kelly et al., 2011) (Yi et al., 2011). Alternatively, nuclear β -catenin may compete with the transcriptional repressor TCF3 to either activate or relieve the repression of pluripotency factors. Given that TCF3 colocalises and antagonises OCT4 and SOX2 at core pluripotency genes, including *Nanog*, *Tfcp2l1* and *Esrrb*, disruption of TCF3-mediated repression would create a feed-forwards loop that stabilised the naive pluripotent state. This is evidenced by the observation that *Tcf3* deletion mimics the effects of GSK3 β inhibition (Wray et al., 2011). Additionally, truncated, transcriptionally inert β -catenin can rescue the effects seen following β -catenin depletion (Kelly et al., 2011), suggesting that β -catenin does not solely depend on transcriptional activation. Studies have also shown that TCF3 becomes rapidly phosphorylated in response to WNT signalling and GSK3 inhibition and β -catenin stabilisation. Therefore, one model proposes that β -catenin may form a scaffold to which both TCF3 and the kinase HIPK2 can bind (Hikasa et al., 2010).

Interestingly, the aforementioned cytokines and inhibitors additively reinforce pluripotency. Although they primarily target different pathways (JAK/STAT, FGF/ERK and WNT), any combination of two of the three supplements is sufficient to maintain mESC self-renewal (Wray et al., 2010)(Dunn et al, 2014), indicating that to some extent, these pathways must be redundant, with a degree of cross talk and some shared targets.

1.7.8 Defining the naive ground state

Pluripotent cells are often described as being naive or ground state. The pluripotent epiblast is functionally and molecularly distinct from the blastomeres and the early ICM (Gardner et al., 1998)(Kaji et al., 2007)(Kurimoto et al, 2007) and represents the true developmental ground state (the tubula rasa). mESCs grown in 2i are also commonly described as being naive or ground state. Although related, these concepts are ultimately distinct. Whilst naivety is a functional description of unrestricted and unbiased developmental potential (i.e. the ability to contribute to chimaeras), the latter reflects the epigenetic and transcriptional status associated with the naive pluripotent cells of the epiblast. Since both conventional FCS and 2i cultured mESCs contribute to somatic and germline tissues in chimaeras, both conditions harbour naive pluripotent mESCs. However, only 2i/LIF

conditions indefinitely capture mESCs with neither developmental nor epigenetic constraints and with a transcriptional profile similar to the E4.5 epiblast (Boroviak et al., 2014). Therefore only mESCs cultured in 2i/LIF conditions are likened to the pluripotent ground state.

Under conventional FCS culture conditions rich in growth factors and cytokines, mESC are subjected to ill-defined and often conflicting signalling pathways; promoting morphological, transcriptional and functional heterogeneity (Kolodziejczyk et al., 2015)(Marks et al., 2012). Thus conventional mESCs exist in at least two distinct populations; one corresponding to a developmentally naive state, and one which more closely resembles the primed state of the post implantation epiblast. These two states are interchangeable in culture. Further populations may exist that are primed towards the extraembryonic fate or that are irreversibly committed to differentiation (Canham et al., 2010).

Crucially, inhibition of GSK3 β and ERK1/2 signaling is thought to shield mESCs from differentiating stimuli and select against differentiating cells, thus capturing mESCs at a point prior to commitment and restriction of cell fate decisions. In line with this theory, 2i cultured mESCs show reduced or near absent expression of lineage-associated genes, a permissive epigenetic landscape and at the single cell transcriptome level, cluster closely with E4.5 epiblast cells (Boroviak et al., 2014)(Marks et al., 2012). In addition to uniform expression of key pluripotency factors, mESCs cultured in 2i conditions also show high expression of *Nanog* mRNA and the transcriptional repressor *Prdm14* (Silva et al., 2009)(Leitch et al., 2013). 2i media also facilitates the derivation of ESCs from biologically important recalcitrant strains including NOD and FBV (Kanda et al., 2012)(Nichols et al., 2009)(ten Berge et al., 2011).

1.7.9 Transcriptional control of pluripotency

1.7.9.1 Core pluripotency factors

Pluripotency factors directly and specifically support the maintenance of mESC identity. At the heart of this transcriptional network is the core triad of OCT4, SOX2 and NANOG. These factors cooperatively bind overlapping targets (including their own genes) to form a robust, self-regulatory transcriptional network.

OCT4, a POU-domain transcription factor, was the first factor to be identified and characterised as a regulator of pluripotency (Nichols et al., 1998). *In vivo*, the *Oct4* gene is expressed in the ovum prior to fertilisation and transcribed in the zygote from the late two cell stage onwards (Scholer et al., 1989). At cleavage, OCT4 is distributed evenly between all cells. However, in the early blastocyst, *Oct4* gene expression is gradually downregulated in the trophectoderm and although transiently expressed in the primitive endoderm, it is silenced in extraembryonic tissue following implantation (Palmieri et al., 1994). Blastocysts develop in the absence of OCT4, however the ICM will inevitably differentiate into trophoblast (Nichols et al., 1998). Therefore, it is likely that OCT4 is essential for the differentiation of the ICM into the epiblast. *Oct4* expression continues into early somitogenesis, after which it is exclusively expressed in the germline. *Ex vivo*, OCT4 is only found in pluripotent EC, ES and EG cells.

Similarly, SOX2 is required for the acquisition and maintenance of pluripotency. Both *Sox2* and *Oct4* are expressed at comparable levels between conventional FCS and 2i culture conditions. In the absence of either factor, mESCs gradually differentiate towards the trophectoderm lineage (Chew et al., 2005)(Niwa et al., 2000). Overexpression also results in lineage commitment, suggesting that their expression must be maintained within precise limits to balance self-renewal and differentiation (Masui et al., 2007)(Niwa et al., 2000)(Thomson et al., 2011).

The third factor in this tripartite network, *Nanog*, is transcribed following activation of zygotic transcription (Guo et al., 2010). Although dispensable for blastocyst formation, NANOG is absolutely required for

differentiation of the ICM into the epiblast (Mitsui et al., 2003). However, unlike OCT4, *ex vivo* NANOG is dispensable once mESC cultures have attained pluripotency. Additionally, NANOG loss does reduce the threshold for differentiation (Chambers et al., 2007). Therefore, it has been suggested that NANOG's primary role is to establishment the ground state of pluripotency, whilst its secondary function is the stabilisation of self renewal in diapause and mESCs (Silva & Smith, 2008).

1.7.9.2 Ancillary pluripotency factors

Ancillary transcription factors, including KLF2, ESRRB, KLF4, PRDM14, SALL4, TFCP2L1, and TBX3, reinforce and buffer the pluripotency network against pro-differentiation stimuli. Typically, ancillary factors stabilise each other's expression by regulatory feedback loops. In conventional FCS culture conditions these factors display heterogeneous expression; fluctuating between on and off states at the single cell level (Martello et al., 2012)(Chambers et al., 2007)(Toyooka et al., 2008)(Yamaji et al., 2013). At the population level, some ancillary factors are expressed at slightly higher (typically less than 2-fold) levels in 2i compared to conventional FCS culture conditions. It is possible that this represents a switch from mosaic to stable expression (Marks et al., 2012). Other genes, including *Prdm14*, *Stella*, *Tcf1* and *Tfcp2l1*, are directly and significantly upregulated by between 2 and 5 fold in cells grown in 2i; suggesting that they may be important for the stabilisation of the ground state. Consistent with this theory, overexpression of *Tfcp2l1* can efficiently reprogram EpiSCs to naive pluripotency (Grabole et al., 2013)(Martello et al., 2013).

1.7.10 Prdm14 and the pluripotency network

The transcriptional repressor PRDM14 links the pluripotency transcriptional network and epigenetic regulators of naive pluripotency. *Prdm14* is specifically and asymmetrically expressed in the preimplantation embryo; with *Prdm14*-expressing blastomeres invariably later forming the pluripotent cells of the ICM (Burton et al, 2013). Upon differentiation, *Prdm14* gene expression ceases, only to be re-expressed following primordial germ cell specification at E6.5 (Yamaji et al., 2008). The expression of *Prdm14* in PGCs is believed to play a role in establishing a latent state of pluripotency (Yamaji et al., 2008). Loss of *Prdm14*, however, is associated with

infertility in mice due to the loss of migratory PGCs (Yamaji et al., 2008). Little or no *Prdm14* is expressed in normal mouse or human somatic tissues. However, aberrant reactivation of *Prdm14* is frequently seen in human malignancies (Nishikawa et al., 2007)(Dettman et al., 2011)(Carofino et al., 2013)(Zhang et al., 2013). These cancer cells also express a series of pluripotency markers, suggesting that *Prdm14* functions as a proto-oncogene; ultimately forming a population of cancer stem cell-like progenitor cells.

The role of PRDM14 in pre-implantation- and PGC-development is not fully understood; largely due to the difficulties associated with working with small numbers of cells. However, PRDM14 is also likely to play an equivalent role in mESCs. Genetic knockout studies have illustrated how the loss of *Prdm14* in mESCs destabilises the pluripotent state by sensitizing the cells to differentiation-inducing FGFR signaling (Yamaji et al., 2013). Loss of *Prdm14* is also accompanied by global DNA hypermethylation (Yamaji et al., 2013)(Leitch et al., 2013)(Hackett et al., 2013)(Grabole et al., 2013). Since gene expression analysis has shown that PRDM14 occupies, and presumably represses, the regulatory elements of genes involved in FGFR signalling (*Fgfr1*, *Fgfr2*, *Fgfbp1*, *Hs6st2*, and *Epha4*) and DNA methylation (*Dnmt3a*, *Dnmt3b*, and *Dnmt3l*), these phenotypes are likely to be a direct result of *Prdm14* ablations (Yamaji et al., 2013)(Magnusdottir et al., 2013).

PRDM14 also directly interacts with PRC2. Therefore, the binding of PRDM14 to target sequences is associated with increased histone methyltransferase activity and, since H3K27me3 is a repressive modification, a repressive chromatin state. This process is known to contribute to the silencing of *Dnmt3a*, *Dnmt3b* and *Dnmt3L* expression.

More recent studies have suggested an alternative, perhaps complementary means of PRDM14-mediated gene repression. Specifically, coimmunoprecipitation studies identified CBFA2T2 (MTGR1), a co-repressor known to co-occupy PRDM14 target loci, as having a strong association with PRDM14 (Tu et al., 2016)(Nady et al., 2015). Deletion of either *Prdm14* or *Cbfa2t2* causes the derepression of *Ehmt1* (*Glp*) H3K9 methyltransferase (Tu et al., 2013). Accordingly, in *Prdm14* null mESCs, *Glp* derepression results in elevated levels of both H3K9me2 and H3K9me3. In contrast, H3K9me2 levels are significantly reduced in *Prdm14*-expressing PGCs and 2i-cultured mESCs (Leitch et al., 2013). These experiments suggest that PRDM14 is a potential master regulator

involved in the regulation of global DNA methylation levels via repression of the Dnmt3 family and the global regulation of histone modifications.

Chapter 2

Materials and Methods

2.1 Mice

2.1.1 Mouse husbandry

All mouse work was carried out in a Home Office designated facility under a UK Home Office project license. Cryopreserved B6.129S4-*Dnmt1*^{tm2}*Jae* spermatozoa was purchased from the Jackson Laboratory. The strain was recovered by the CSC Transgenics Facility. Homozygous *Dnmt1* F/F mice were bred onto a homozygous *GOF18ΔPE-EGFP* (Yoshimizu et al 1999) line. *Prdm1*-Cre was used to drive the germline deletion of *Dnmt1*. Specifically, *B6.Cg-Tg(Prdm1-Cre)* mice (Ohinata et al., 2005) were crossed with homozygous *Dnmt1* F/F mice. Male offspring with the genotype *Dnmt1* Δ/+ *Prdm1-Cre*, were then crossed with female *GOF18ΔPE-EGFP Dnmt1* F/F to produce *Dnmt1* knockout PGCs.

2.1.2 Genotyping

Tissue samples were collected from mice by ear punching. Samples were genotyped by overnight digestion at 55°C in genotyping buffer (50 mM KCl, 10 mM Tris-HCl, 0.1 mg/ml gelatin, 0.45% Tween 20 in H₂O) supplemented with 500 µg/ml Proteinase K (Qiagen). Proteinase K was heat inactivated by incubation at 95°C for 10 min. Crude DNA extracts were used directly for PCR using RedTaq Ready Mix and the primers and conditions detailed below:

Table 2.1: Mouse genotyping primers and PCR protocols

Gene	Forwards primer	Reverse primer
<i>Dnmt1</i> F/F	GGGCCAGTTGTGTGACTTGG	CTTGGGCCTGGATCTTGGGGA
0.4 µM each primer & 1.5 µl DNA extract in a total volume of 20 µl. 1. Initiation: 94°C , 5 minutes. 2. Denaturation: 94°C , 15 seconds 3. Annealing: 65-55°C, 30 seconds 4. Elongation: 72°C , 40 seconds Repeat steps (2-4) 10 times, reducing the annealing temperature by 1°C each cycle. 5. Denaturation: 94°C , 15 seconds 6. Annealing 55°C , 30 seconds		

7. Elongation: 72°C , 40 seconds Repeat steps (5-7) 30 times. 8. Amplification: 72°C, 5 minutes 9. Finish, 4°C, ∞		
<i>Prdm1-Cre</i>	GCCGAGGTGCGCGTCAGTA	CTGAACATGTCCATCAGGTTCTTG
0.4 μM each primer & 1.5 μl DNA extract in a total volume of 15 μl. 1. Initiation: 94°C , 5 minutes. 2. Denaturation: 94°C , 15 seconds 3. Annealing: 60°C, 30 seconds 4. Elongation: 72°C , 40 seconds Repeat steps (2-4) 39 times. 5. Amplification: 72°C, 5 minutes 6. Finish, 4°C, ∞		
<i>Tnap-Cre</i>	TGATGAGGTTGCAAGAACC	CCATGAGTGAACGAACCTGG
0.4 μM each primer & 1.5 μl DNA extract in a total volume of 15 μl. 1. Initiation: 94°C , 5 minutes. 2. Denaturation: 94°C , 15 seconds 3. Annealing: 59°C, 30 seconds 4. Elongation: 72°C , 40 seconds Repeat steps (2-4) 39 times. 5. Amplification: 72°C, 5 minutes. 6. Finish, 4°C, ∞		
<i>ΔPE Oct4 GFP</i>	GGTGAACCTCAAGATCCGCC	CTTGTACAGCTCGTCCATGC
0.4 μM each primer & 1.5 μl DNA extract in a total volume of 15 μl. 1. Initiation: 94°C , 5 minutes. 2. Denaturation: 94°C , 15 seconds 3. Annealing: 58°C, 30 seconds 4. Elongation: 72°C , 40 seconds Repeat steps (2-4) 39 times. 5. Amplification: 72°C, 5 minutes. 6. Finish, 4°C, ∞		

2.2 mESC culture protocols

2.2.1 Culture conditions

FCS medium consists of DMEM-F12 (Gibco) supplemented with 15% FCS, 0.1 mM MEM nonessential amino acids, 2 mM L-glutamine, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol and ESGRO mouse Leukemia inhibitory factor (LIF) 10 µg/ml (EMD Millipore, Cat. no. ESG1107). Unless stated otherwise, cells cultured in serum were maintained on a mouse embryonic fibroblast (MEF) feeder layer. For downstream analysis, MEFs were either removed by serial panning or fluorescent activated cell sorting (e.g. using Anti SSEA1- eFluor 660, eBiosciences, 1:20 dilution).

2i medium consists of the MEK inhibitor PD0325901 (PD) at 1 µM, the GSK3 inhibitor CHIR99021 (CH) at 3 µM and mouse LIF at 10 µg/ml in N2B27 medium. N2B27 medium consists of a 1:1 ratio of DMEM-F12 (Gibco) and Neurobasal (Gibco) supplemented with N2 (Invitrogen) and B27 (with or without Vitamin A, as stated)(Invitrogen), 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol and mouse LIF 10 µg/ml.

¹⁵N₂-Uridine and ¹³C-d₃ L-methionine medium consists of DMEM (without cysteine or L-methionine), supplemented with 0.2 mM L-cysteine 2HCl, 100 µM ¹⁵N₂-Uridine (Sigma, 7939809), 0.2 mM ¹³C-d₃ L-methionine (Sigma, 793809), 15% dialysed FCS (Biowest), 0.1 mM MEM nonessential amino acids, 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol and ESGRO mouse Leukemia inhibitory factor (LIF) 10 µg/ml.

KSR culture medium consists of KnockOut DMEM (KO DMEM)(Gibco) supplemented with 15% Knockout Serum Replacement (KSR), 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol and mouse LIF 10 µg/ml.

FCS medium + 2i consists of FCS medium (as above) supplemented with 1 µM PD0325901 and 3 µM CHIR99021.

MEF medium consists of FCS medium (as above) minus 10 µg/ml LIF.

2.2.2 mESC lines

Details of the mESC lines used during the studies described in this thesis are outlined below:

Table 2.2 Mouse embryonic stem cell lines

Name	Background	Source	Passage No. used during these studies
E14	129/Ola	MRC CSC Transgenic Facility	18 - 23
12g4	129/GGOF	Dr. H. Leitch	15 - 22
12g3	129/GGOF	Dr. H. Leitch	16 - 18
Tet1/Tet2 DKO	C57B6	Prof. R Jaenisch	19 - 25
<i>Tet1/Tet2</i> WT	C57B6	Prof. R Jaenisch	19 - 25
<i>Dicer</i> KO	C57B6	Prof. N. Brockdorff	17 - 26
<i>Dnmt1</i> F/F	C57B6	Dr. H. Leitch & S. Linnett	17 - 26

2.2.3 Passaging

All mESCs were incubated in a humidified atmosphere at 37°C with 5% CO₂. Media was replaced daily and 2 hours prior to passaging. To passage mESCs, the culture media was aspirated and cell colonies were washed once in pre-warmed PBS. Enough TrypLE Express enzymatic dissociation reagent (ThermoFisher, Cat no. 12604013) was added to cover the growth area. Cells were incubated with the dissociation reagent for 3 minutes at room temperature. The enzymatic dissociation reagent was quenched by the addition of an equal volume of culture medium. The mESCs were then dislodged by gentle agitation and collected by centrifugation. The cell number was counted and a single cell suspension was seeded on either gelatin coated (0.2% in PBS) Nunclon Delta treated plastic dishes or mitomycin C treated MEF monolayer at a concentration of 1.5×10^5 cells ml⁻¹.

2.2.4 Cryopreservation and thawing

Cell suspensions in FCS culture media supplemented with 10% (v/v) dimethyl sulphoxide (DMSO) were frozen in liquid nitrogen. To reduce loss of cell viability, mESCs were added to FCS medium + 20% DMSO in a 1:1 ratio. Typically, the final mESC concentration was 1×10^6 cells ml^{-1} . mESCs maintained in FCS-free conditions were resuspended directly in serum-free freezing medium (Gibco) at a concentration of 1×10^6 cells ml^{-1} . The rate of freezing was controlled ($-1^\circ\text{C min}^{-1}$). For long-term storage, cryovials were placed in the gas-phase of a liquid nitrogen tank. Cells were retrieved from liquid nitrogen by thawing in a 37°C water bath, followed by dropwise dilution of the cell suspension in 10 ml pre-warmed (37°C) culture medium. The cells were collected by centrifugation, to remove the DMSO, and seeded on a 10 cm Nuclon treated culture dish.

2.2.5 Cell counting

Cells were harvested into a single cell suspension in PBS at an approximate concentration of 1×10^6 cells ml^{-1} for counting. 0.4% Trypan blue solution diluted 1:5 in an aliquot of cell suspension. The cell number were calculated using a Neubauer haemocytometer (0.1 mm depth, $1/400 \text{ mm}^2$) and the following equation:

$$\text{Concentration (cells ml}^{-1}\text{)} = \frac{(\text{Number of cells counted} \times 10^4)}{(\text{Number of large squares of volume } 10^{-4} \text{ ml} \times \text{dilution factor})}$$

2.2.6 mESC derivation

Timed pregnancies between *Dnmt1 F/F ΔPE-Oct4 GFP* mice were set up and females were checked daily for the formation of a vaginal plug. Pregnant females were euthanised by cervical dislocation 3.5 days post fertilisation. The peritoneum was opened by a lateral incision and the uterine horns, oviducts and ovaries lying beneath the intestines were exposed. Two incisions were made in the uterus: firstly between the the oviduct and the uterine horn and secondly in the uterus body (uterine fundus) adjacent to the cervical region. The uterine horns were flushed with M2 medium using a 27 gauge (27-G) needle. Using a mouth pipette, embryos were transferred to KSOM medium and cultured until a discernable blastocoel could be seen (12 - 36 hours). During this time, the

zona pellucida began to visibly disintegrate. Single late-stage blastocysts were then transferred to individual MEF-coated wells where they were cultured in 2i medium for approximately 10 days; until a clear pluripotent outgrowth could be seen. Outgrowths were scored for the expression of *ΔPE-Oct4 GFP* using a brightfield microscope. The outgrowths were enzymatically disaggregated using 0.25% trypsin. Trypsin was completely quenched using FCS-containing mESC media. The mESCs were subsequently passaged in FCS-free and feeder-free conditions on laminin (10 µg/ml). Stocks of low-passage cells (<p6) were cryopreserved for future experiments.

2.2.7 ESC transfection

2.2.7.1 Nucleofection

Fresh 2i medium was added to *Dnmt1 F/F* female mESCs 4 hours prior to transfection. mESCs were transfected using the mouse ES Cell Nucleofector kit (Lonza) according to the manufacturer's protocol. Briefly 0.5 ml Nucleofector supplement was added to 2.25 ml Nucleofector solution and the supplemented Nucleofector reagent prewarmed to room temperature. Cells were harvested as above using TrypLE Express. 10 ml 2i culture medium (supplemented with 100 U/ml Penicillin Streptomycin) was prewarmed for every transfection. Cells were washed in TrypLE wash medium (DMEM F-12, 0.1% BSA) and pelleted by centrifugation at 80 x g for 5 min at 4°C. The pellet was washed in PBS and centrifuged as before. Whilst pelleting the cells, 10 µg linearised DNA was diluted with supplemented Nucleofector reagent to a final volume of 10 µl. 5×10^6 cells were resuspended in 90 µl ESC Nucleofector solution. 90 µl cell suspension was added to 10 µl linearised DNA and mix by gentle pipetting. The nucleofector solution was transferred to an Amexa certified cuvette. The pre-programmed A-24/A-024 mESC specific programme was chosen to electroporate the cells. 500 µl prewarmed culture medium was added to the cuvette. Electroporated cells were carefully remove from the cuvette and immediately plated onto a laminin-coated (10 µg ml⁻¹ in PBS) 10 cm culture dish containing pre-warmed medium.

2.2.7.2 Puromycin selection

24 hours post-transfection, 2i culture medium was replaced with 2i medium supplemented with 100 U/ml Penicillin Streptomycin and 1 µg/ml puromycin. Fresh medium was added to each plate 3 and 5 days after the addition of puromycin. 7 days post-transfection, well-formed spherical colonies were picked under a dissecting microscope. Colonies were washed in droplets of PBS and dissociated in droplets of TrypLE Express. Trypsin was quenched by the addition of DMEM F-12 + 5% BSA (Fraction V, Gibco). The colonies were transferred to individual droplets of 2i medium and disaggregated by pipetting. The single cell suspensions were then plated in prewarmed 2i medium supplemented with 100 U/ml Penicillin Streptomycin and 1µg/ml puromycin on laminin-coated (10 µg ml⁻¹) coated plates. Unless otherwise stated, for further passaging puromycin was removed.

2.2.7.3 Optimising Puromycin concentration using a kill curve

5 x 10⁵ non-transfected *Dnmt1* F/F cells were seeded in 33 mm Nuclon treated dishes. After 24 hours, 2i medium supplemented with either 0.5, 1.5, 2, 2.5 or 3 µg ml⁻¹ puromycin was added. The optimal concentration of puromycin was determined as the minimum concentration required to effectively kill all non-transfected cells within 48 hours.

2.2.7.4 Tamoxifen treatment

Single cell suspensions of *Dnmt1* F/F mESCs were seeded onto gelatinised plates (0.2%) in 2i medium supplemented with 250 nM, 500 nM, 750 nM, 1000 nM and 1250 nM 4'hydroxytamoxifen (4'OHT). Spent culture medium was replaced daily. Cell pellets were collected after 24, 48, 72 and 96 hours for PCR, RT-qPCR and western blot analysis of *Dnmt1* expression. This procedure was also repeated in the presence of 1 µg ml⁻¹ puromycin.

2.2.8 Preparation of feeder cells

2.2.8.1 MEF derivation

The head, limbs and internal organs were removed from the abdominal cavity of embryonic day (E) 13.5 embryos. The embryonic tissue was mechanically sheared using a 16-G needle. 0.25% trypsin/EDTA was added and the tissue incubated at 37°C for 5 minutes. A single cell suspension was created by gentle pipetting. Care was taken not to over-trypsinise or over-dissociate the cells. Trypsin was inactivated by the addition of an equal volume of FCS-containing medium. Large pieces of intact tissue were allowed to settle to the bottom of the tube. Cells from the supernatant were then pelleted, resuspended in fresh medium and seeded on a gelatinised (0.2% in PBS) plate. The number of cells harvested from 2 embryos is sufficient to seed a 15 cm dish. Cells were passaged twice and frozen at passage 3. P3 stocks were thawed and expanded for a further 3 passages. After reaching confluency, MEFs were split by a ratio of no more than 1:4.

2.2.8.2 MEF inactivation

At p7, MEFs were inactivated by treatment with mitomycin C (MMC). MMC was resuspended in 10 ml PBS at a concentration of 10 $\mu\text{g ml}^{-1}$. 500 μl stock MMC was diluted 1:20 (final concentration 0.5 $\mu\text{g ml}^{-1}$) in MEF medium and added to the cells. Cells were incubated in MMC for 3 hours in a 37°C humidified incubator. After MMC treatment, cells were thoroughly washed 3 times with prewarmed PBS, dissociated and frozen as stocks of 5×10^6 cells.

2.3 Differentiation protocol

2.3.1 Myoblast differentiation

2.3.1.1 C2C12 medium

C2C12 expansion medium consists of high glucose DMEM supplemented with 10% foetal bovine serum (FCS, Source Bioscience, EU-000-F), 10% horse serum (HS, Gibco, 16050-122) and 2 mM L-glutamine (Gibco).

C2C12 differentiation medium consists of high glucose DMEM (PAA, E15-011), supplemented with 2% horse serum (HS) and 2 mM L-glutamine (Gibco).

2.3.1.2 C2C12 differentiation

1 x 10⁶ myoblasts were seeded per 10 cm dish. The myoblasts were grown in expansion medium until they reached 70%–80% confluency. Differentiation was induced through mitogen withdrawal by replacing the FCS-rich expansion medium for differentiation medium containing 2% horse serum (see below). Differentiation was assessed based on morphological changes (i.e. the formation of syncytial myotubes).

2.3.1.3 Dialysis of horse serum for stable isotope labelling

Prior to use in our stable isotope labelling experiments (Chapter 4), horse serum was dialysed using SnakeSkin Dialysis Membrane (Thermo) with a 3.5K molecular weight cutoff (MWCO). Dialysis tubing containing the horse serum was sealed with a weighted clip at one end, and a bouyant clip at the other. 4 litres of sterile 1X PBS was added to an autoclaved beaker. The dialysis tubing was immersed in the PBS and stirred constantly at 4°C. After 6 hours, the spent PBS was replaced and the dialysis membrane stirred overnight. The following day, the spent PBS was again replaced and the dialysis membrane stirred for another 6 hour. Dialysed horse serum was filter sterilised (0.22 µm) prior to storage at -80°C .

2.3.2 Adipogenesis

2.3.2.1 Composition of 3T3-L1 expansion medium and adipogenic cocktail

3T3-L1 expansion medium consists of high glucose DMEM (Gibco) supplemented with 10% FCS, 100 U ml⁻¹ Penicillin Streptomycin and 4 mM L-glutamine.

3T3-L1 adipogenic induction medium 1 consists of 3T3-L1 expansion medium supplemented with 500 µM IBMX, 1 µM Dexamethasone, 1 µM Thiazolidinedione (TZD) and 5 µg ml⁻¹ insulin.

3T3-L1 adipogenic induction medium 2 consists of 3T3-L1 expansion medium supplemented with 1 µM Thiazolidinedione (TZD) and 5 µg ml⁻¹ insulin.

3T3-L1 adipogenic induction medium 3 consists of 3T3-L1 expansion medium supplemented with 0.5 µg ml⁻¹ insulin.

2.3.2.2. 3T3-L1 adipogenic differentiation protocol

3T3-L1 cells were cultured on gelatinised plates (0.1%) in expansion medium until they reached 70% confluency. Spent medium was replaced by half medium changes every 2 days. To initiate differentiation, 3T3-L1 expansion medium was replaced with 3T3-L1 adipogenic induction medium 1. 4 days post adipogenic induction, induction medium 1 was replaced with induction medium 2. To complete the adipogenic induction, after 2 days, induction medium 2 was replaced with induction medium 3. Throughout the induction process, 3T3-L1 cells were incubated in a humidified atmosphere at 37°C with 5% CO₂.

2.3.2.3 Oil Red O staining

To assess the proportion of mature adipocytes following adipogenic induction, cells were stained with Oil Red O. A stock solution of Oil Red O was prepared by dissolving 300 mg Oil Red O powder in 100 ml isopropanol (7.35 M). A working solution of Oil Red O was then prepared by adding 3 parts Oil Red O to 2 parts ddH₂O. Before use, the working solution was filtered to remove particulate matter. Differentiated 3T3-L1 cells were rinsed with PBS prewarmed to 37°C and fixed with 4% PFA for 30 minutes at room temperature. Cells were washed twice using prewarmed PBS and were then incubated with 60% isopropanol for 5 minutes at room temperature. Finally, the cells were stained with the Oil Red O working solution at for 5 minutes at room temperature. Before imaging, the cells were carefully washed with prewarmed PBS to remove background cytoplasmic staining.

2.3.3 Bone marrow derived macrophage protocols

2.3.3.1 Culture medium

L929-conditioned medium was prepared by plating 5×10^5 L929 cells in a T175 with 50 ml L929 medium. Cells were grown in a humidified incubator with 5% CO₂ at 37°C until they reached confluency (approximately 5 days). During this time, spent medium was *not* replaced. The supernatant was collected, sterilised using a 0.45 µm filter and stored in 25 ml aliquots at -80°C .

BMΦ medium consists of DMEM (high glucose 4.5 g/L)(Gibco), 25 mM Hepes, 2 mM L-glutamine, 100 U ml⁻¹ Penicillin Streptomycin, 25% FCS and 25% L929-conditioned medium.

2.3.3.2 BMΦ derivation

Sterile technique was used throughout the derivation procedure. All utensils were sterilised using 70% ethanol and when possible, a laminar flow hood was used. 6 week old male C57B6 mice were euthanised by cervical dislocation. 70% ethanol was used to sterilise the abdomen and hind legs. A midline incision was made to

exposed the hindlegs and allow the muscle to be separated from the femurs using dissection scissors. Subsequent removal of the the patella facilitated the separation of the femur and tibia. By severing the femur at the hip joint, the intact femur was removed. The process was repeated to retrieve the remaining femur. If higher cell yields were required, the tibias were also removed. Excess muscle and cartilage was removed using a scalpel before briefly washing the exterior of the bone with 70% ethanol. A sharp pair of scissors were used to sever the proximal and distal ends of the bone. A 25-G needle attached to a syringe containing 10 ml ice-cold Hanks buffered salt solution (HBSS) was inserted into the honeycomb-like trabecular tissue at one end of the bone. *Slowly* 5 ml of medium was forced through the bone, and the bone marrow filtrate collected in a sterile 50 ml tube. The process was repeated from the opposite end of the femur until the bone appeared brilliant white. Extracts were incubated at 37°C for 30 minutes to lyse any contaminating erythrocytes. The cells were pelleted for 10 minutes at 500 x g at room temperature and gently resuspended in BMΦ medium prewarmed to 37°C . The cell suspension was filtered through a 70 μM filter to remove remaining intact tissue prior to seeding on untreated sterile petri-dishes in complete BMΦ medium. BMΦ culture medium was not changed for the first 3 - 5 days depending on cell confluency.

2.3.3.3 Dissociation of BMΦ

BMΦ were washed once with prewarmed PBS. To dissociate the BMΦ, the cells wer incubated with ice-cold (1X) Non-enzymatic cell dissociation solution (Sigma, Cat. no. C5789) for 10 minutes. Ice cold PBS was added to dilute the dissociation solution. Cells were collected with minimal pipetting and pelleted by centrifugation at 200 x g for 5 minutes at room temperature. Cells were subsequently plated onto non-coated plastic sterile petri dishes.

2.3.3.4 Flow cytometric analysis of BMΦ surface markers

BMΦ were harvested (as above) and washed once in FACS buffer (0.1% BSA (wt/v) in PBS). Cells were pelleted at 200 x g for 5 minutes at room temperature and resuspended in FACS buffer + primary antibody as specified in Table 3. Cell suspensions were incubated on ice protected from the light for 15 min. Subsequently cells were

pelleted and washed once in 5 ml FACS buffer. Cells were pelleted, resuspended in 500 μ l FACS buffer and submitted to the MRC flow cytometry facility for analysis.

Table 2.3: Antibodies used against BM Φ cell surface proteins

Antibody	Source	Dilution (for 10^7 cells ml^{-1})
anti-F4/80	Santa Cruz (sc-377009)	1:50
anti-CD11b PE-Cy7	Abcam (ab25506)	1:300
Anti-GR1 BV605	BD Biosciences (563299)	1:100

2.4 Dissection techniques

2.4.1 Cortical neuron harvest

For the harvest of embryonic cortical neurons for *ex vivo* culture, timed pregnancies were set up between C57B6 mice. 15 days post fertilisation, glass coverslips were sterilised using 70% (v/v) ethanol and incubated with sterile poly-lysine overnight at room temperature. At E16.5 the pregnant females were euthanised by cervical dislocation and the prenatal pups removed from the abdominal cavity. Under a dissecting microscope, the head was removed from each pup using a scalpel. Using dissecting scissors, the cranium was opened from the back of the head to the tip of the nose. The brain was removed using a small, flat angled spatula. The cerebellum was initially removed and the brain separated into its two hemispheres using a pair of sterile needles; one needle was used to anchor the brain whilst the other was used to make the incisions. The meninges was removed using forceps to peel it away from the outer surface. Holding one hemisphere so that the exterior surface was lying against the dish and the interior surface was visible, the midbrain was cut from the cortex using short strokes. The hippocampus was removed using dissecting forceps and the cortices cut into 5-6 pieces, aspirated and transferred to a 15 ml tube. Trypsin (10X) was added to the tissue and incubated for 30 minutes at 37°C. 50 μ l 10 mg ml^{-1} DNase I was added to 10 ml tissue suspension to preventing cell clumping caused by the DNA released during cell lysis. Large pieces of tissue were allowed to settle at the bottom of the tube and, after 5 min,

the supernatant was removed and washed with 10 ml DMEM/10% FCS. The wash was repeated twice and the cell pellet resuspended in 1 ml DMEM/10% FCS. Glass pipettes were fire-polish by heating the tip on a roaring flame. Cells were triturated approximately 10 times to obtain a single cell suspension. Utmost care was taken to avoid damaging the cells. The cell number was calculated using a haemocytometer and trypan blue staining. 9×10^4 cells were plated onto a single coverslip in cortical neuron culture medium.

For cortical neuron harvest, adult male C57B6 mice were fed a custom L-methionine-free diet supplemented with L-methionine-(*methyl*- ^{13}C ,d3) (Sigma, manufactured by TestDiet)(Amouroux et al., 2016) for 5, 7 and 9 weeks respectively. Mice were then euthanised by cervical dislocation. Essentially the method used to isolate embryonic cerebral cortices was used to isolate the adult cerebral cortices. Nuclei were then isolated from the cortices for downstream processing.

2.4.2 Nuclei isolation

Tissue harvested from the adult mouse cortex was Dounce homogenised in PBS using a loose fitting pestle. Cells were pelleted by centrifugation and resuspended in hypotonic lysis buffer (0.32 M sucrose, 5 mM CaCl_2 , 5 mM Mg acetate, 0.1 mM EDTA, 10 mM Tris HCl pH 8.0, 0.1% Triton X-100, 1 mM DTT) and allowed to swell on ice for 10 minutes. Cells were centrifuged at $300 \times g$ for 5 minutes at 4°C and resuspended in lysis buffer. 20 - 30 strokes of a tight fitting pestle were needed to extract the nuclei. The nuclei were then pelleted at $4000 \times g$ at 4°C for 5 minutes in a tabletop centrifuge. The pellet was resuspended with 5 strokes of the tight fitting pestle in nuclei wash buffer (1.7 M sucrose, 5 mM Mg acetate, 1 mM DTT and 10 mM Tris HCl pH 8). Cells were washed twice with nuclei wash buffer. Centrifugation speed was optimised to reduce nuclei clumping and the centrifugation time was optimised to ensure minimal loss of nuclei in the supernatant.

2.4.3 Immunolabelling of nuclei

Isolated nuclei were incubated with nuclei labelling buffer (1% BSA in nuclei wash buffer) for 30 minutes at 4°C to reduce non-specific binding. The primary antibody, anti-NeuN (ab117487), was diluted in nuclei labelling buffer 1:100. Nuclei were pelleted by centrifugation at 4,500 x g at 4°C for 5 minutes and resuspended in the nuclei labelling buffer and primary antibody. Nuclei were incubated with the primary antibody for 30 minutes at 4°C. The secondary antibody was added to the nuclei and incubated for a further 20 minutes. Nuclei were pelleted by centrifugation at 4,500 x g at 4°C for 5 minutes and resuspended in the nuclei wash buffer. This step was repeated twice. Nuclei were resuspended in 500 µl nuclei wash buffer and DAPI was added to the nuclei at a concentration of 3 µM. Nuclei were filtered through a 0.22 µm mesh before FANs analysis. The nuclei were analysed using a FACS Aria II flow cytometer equipped with FACs Diva software.

2.4.4 Gonad dissection

For the study of *Dnmt1* KO PGCs, timed pregnancies were set up between homozygous *GOF18ΔPE-EGFP Dnmt1* *F/F* females and *GOF18ΔPE-EGFP Dnmt1* *F/+ Prdm1-Cre* males. The pregnant females were euthenised by cervical dislocation between 9.5 and 14.5 days post coitum (d.p.c). The abdomen was exposed by making a lateral incision with surgical scissors. The abdominal cavity was opened by making a second lateral incision in the abdominal muscle. Care was taken not to damage the organs beneath by lifting the muscle with forceps before making the incision. If pregnant, the embryos were clearly visible. The uterus was cut at the cervix and the oviducts and transferred to ice-cold PBS. Forceps were used to create a tear in the uterine wall large enough to expose the amniotic membranes and the embryo. The embryos were removed from each conceptus; taking care to ensure that the embryos were not damaged. Each embryo was placed into a separate drop of dissection medium (DMEM, 10% FCS). The heads were removed from each embryo in accordance of UK Home Office regulations. For genotyping, tail tissue was placed in a PCR-clean 1.5 ml tube containing 50 µl genotyping buffer. Embryos were transferred to a clean drop of dissection medium. The embryonic tissue harvested for immunostaining is dependant on the embryonic stage; the embryonic trunk (E9.5 - E10.5), the genital ridge (E11.5 - E12.5) and the gonads (E13.5 - E14.5). In brief, to remove the genital ridge or gonads, embryos were

positions to expose the organs within their abdominal cavity. The liver and intestines were removed to expose the urogenital complex. The genital ridges were separated from the underlying tissue using the forceps to pry the membranes apart. Embryonic tissue was stored in ice-cold PBS for immediate processing. For E12.5 embryos or later, the sex of the embryos was determined based on the presence of visible testis cords.

2.5 Histology

2.5.1 Immunostaining of cryosections

The embryonic tissue was fixed in 2% PFA for 30 minutes at 4°C. Following fixation, the tissue was washed three times with PBS for 10 minutes each. The tissue was then cryoprotected in sterile filtered 30% sucrose overnight with constant agitation. The following day, the tissue was embedded in OCT embedding matrix (Thermo Scientific Raymond Lamb) and flash frozen in liquid nitrogen. Samples were stored at -80°C for future processing. A Leica CM 1950 cryostat was used to cut the tissue into 10 µm sections. The sections were positioned onto a Polylysine-coated glass slide and post-fixed with 2% PFA for a further 30 seconds. The slides were washed three times with PBS for 5 minutes.

Post-fixed sections were incubated with staining buffer (1% BSA, 0.1% Triton X-100 in PBS) for 1 hour at room temperature. Primary antibodies were diluted in staining buffer and incubated with the sections overnight at 4°C. Sections were washed three times with staining buffer for 5 mins and then incubated with a secondary fluorophore-conjugated antibody for 1 hour at room temperature, protected from the light. Excess secondary antibody was removed by three washes with PBS, each for 5 minutes. DNA was then stained with DAPI (0.02 µg ml⁻¹, Severn Biotech) for 10 minutes. Before mounting with Vectashield HardSet (Vector Laboratories), the sections were washed once with PBS.

2.5.2 Immunostaining nucleic acids in cryosections

For 5mdC and 5hmC staining, post-fixed sections were permeabilised with 1% BSA, 0.5% Triton X-100 in PBS for 30 minutes. Following permeabilisation, sections were treated with RNase A (10 mg ml⁻¹; Roche) in 1% BSA in PBS for 1 hour at 37°C. Following three washes in PBS for 5 minutes, genomic DNA was denatured by incubating the sections with 4N HCl for 15, 20 and 25 minutes at 37°C. The sections were washed three times with PBS for 10 minutes and then incubated with blocking buffer (1% BSA, 0.1% Triton X-100 in PBS) for 1 hour at room temperature. Incubation with the primary antibody was carried out at 4°C overnight. Sections were washed with PBS before subsequent secondary antibody staining with the same buffer for 1 hour at room temperature. After three washes with PBS, DNA was stained with propidium iodide (0.25 mg ml⁻¹). After removing excess propidium iodide with a final 10 minute PBS wash, sections were mounted with Vectashield HardSet.

Table 2.4: Immunostaining of cryosections

Antibody	Source	Dilution	Typical secondary antibody
anti-DAZL	Abcam 34139	1:300	Alexa fluor 488 Donkey Anti-Goat IgG (Invitrogen, A11055)
Anti- DDX4	Abcam 27591	1:100	Alexa fluor 488 Goat Anti-Mouse IgG (Invitrogen, A11001)
Anti-OCT4	BD Biosciences 611202	1:200	Alexa fluor 568 Goat Anti-Rabbit IgG (Invitrogen, A11036)
Anti-SYCP3	Abcam 15093	1:300	Alexa fluor 568 Goat Anti-Rabbit IgG (Invitrogen, A11036)
Anti-DNMT1	Abcam 87654	1:5000	Alexa fluor 568 Goat Anti-Rabbit IgG (Invitrogen, A11036)
Anti-5mdC	Diagenode, Clone 33D3	1:10,000	Alexa fluor 405 Goat Anti-Mouse IgG A31553
Anti-5hmdC	Active motif 39791	1:400	Alexa fluor 488 Goat Anti-Rabbit IgG (Invitrogen, A11008)

2.5.3 Haematoxylin and Eosin staining

For testis H&E staining, 4 week old male *GOF18 Δ PE-EGFP Dnmt1 Δ /F Prdm1-Cre* and *GOF18 Δ PE-EGFP Dnmt1 F/+* littermates were euthanised by cervical dislocation. The testes were collected and any remaining blood vessels, nerves and adipose tissue removed under a dissecting microscope. For ovarian H&E staining, 6 week old female *GOF18 Δ PE-EGFP Dnmt1 Δ /F Prdm1-Cre* and *GOF18 Δ PE-EGFP Dnmt1 F/+* littermates were euthanised by cervical dislocation. The ovaries were collected and the surrounding adipose tissue removed under a dissecting microscope. Both testes and ovaries were fixed in 4% PFA for 48 hours at 4°C. After 2 days, the organs were cryoprotected in 30% sucrose for 4 hours at 4°C, embedded in OCT Embedding Matrix (Thermo Scientific Raymond Lamb) and were flash frozen in liquid nitrogen and stored at -80°C for further processing.

10 μ m sections were cut from the frozen embedded tissue using a Leica CM 1950 cryostat. Sections were thaw mounted onto polylysine coated glass slides and allowed to air dry for no longer than 5 minutes. Slides were stained with filtered 0.1% Mayers Haematoxylin for 10 minutes. Excess haematoxylin was removed by rinsing the slides in cool running tap water for 4 minutes followed by 1 minute blueing in Scott's Tap Water (0.25 M Sodium Bicarbonate, 30 mM Magnesium Sulphate in ddH₂O). Excess Scott's Tap Water was removed before staining the slides in 0.5% Alcoholic Eosin. Testes sections were stained for 5 minutes whilst ovary sections were stained for 3 minutes. Excess Eosin was removed by briefly rinsing in 95% ethanol. Slides were dehydrated using 100% ethanol and cleared for 2 minutes in Xylene. Coverslips were mounted using DPX xylene based mountant (Fluka) before imaging.

2.6 Immunocytology

2.6.1 Preparation of glass coverslips

12 mm round glass coverslips were incubated in 1M HCl for 8 hours at 55°C. After cooling to room temperature, the HCl was aspirated and the coverslips washed twice for 5 minutes in sterile ddH₂O. The coverslips were then added to individual wells of a 12 well plate and immersed in 70% ethanol for 5 minutes.

This step was repeated twice before washing the coverslips with 100% ethanol. The 100% ethanol was aspirated and the coverslips rinsed with sterile double distilled H₂O. Finally, coverslips were exposed to UV light for 30 minutes. Before seeding mESCs, coverslips were coated with 10 µg ml⁻¹ laminin in PBS for 4 hours. Laminin was aspirated and washed 3 times with sterile PBS.

2.6.2 Immunofluorescent labelling of cultured cells

Cells were seeded onto acid-washed, sterilised glass coverslips in the appropriate culture condition at a density of 0.1×10^6 . Cells were cultured for 48 hours and were then fixed in 2% PFA for 20 minutes at room temperature. Cells were blocked and permeabilised by incubation with staining buffer (10% horse serum, 0.3% Triton X-100 in PBS) for 1 hour at room temperature. Cells were incubated with their primary antibody overnight at 4°C. The coverslips were washed three times with staining buffer for 5 minutes. Coverslips were then incubated with the secondary antibody for 1 hour at room temperature. Cells were washed once with staining buffer for 5 minutes and twice with PBS. Finally, the cells were stained with DAPI (0.02 µg ml⁻¹, Severn Biotech) for 10 minutes. Before mounting with Vectashield HardSet (Vector Laboratories), the sections were washed once more with PBS.

2.6.3 EdU labelling of adherent mESCs for colocalisation studies

12g4 mESCs were cultured on acid washed, sterilised coverslips. After 48 hours, half the culture medium was replaced with prewarmed medium supplemented with 20 µg ml⁻¹ EdU. Cells were incubated in a humidified atmosphere at 37°C. After 10 minutes, the cells were washed with PBS and fixed with 2% PFA at room temperature for 15 minutes. The PFA was aspirated and cells washed twice with prewarmed PBS. Cells were permeabilised in staining buffer (10% horse serum, 0.3% Triton X-100 in PBS) overnight at 4°C. Covalent addition of the azide conjugated fluorophore was performed according to the manufacturer's protocol. Briefly, for each reaction, a cocktail consisting of 438 µl PBS, 50 µl 1X buffer additive, 10 µl 100 mM aqueous CuSO₄, and 2.5 µl Alexa Fluor 647 azide was incubated with each coverslip for 30 minutes at room temperature (protected from the light). Additional staining for intracellular proteins was then performed as described above.

2.7 Flow cytometry

2.7.1 EdU labelling for flow cytometric studies

Cells were labelled with 10 $\mu\text{g ml}^{-1}$ EdU for either 15 minutes (mESCs) or 2 hours (C2C12, BM Φ). The cells harvested using TrypLE Express and washed in FACS buffer (1% BSA Fraction V, PBS). The cells were fixed in 2% PFA for 15 minutes at room temperature. After washing twice with FACS buffer, cells were resuspended in Saponin-based permeabilisation reagent (Thermo). The cells were permeabilised for 45 minutes at room temperature prior to the covalent addition of the fluorophore conjugated azide (as described in “EdU labelling of adherent mESCs”). Cells were counter-stained using DAPI to aid doublet discrimination and for cell cycle analysis and submitted to the MRC flow cytometry facility for analysis.

2.7.2 Flow cytometry of nuclear proteins

Cells were washed with pre-warmed PBS and harvested the using TrypLE Express. Either FCS medium or DMEM + BSA was used to quench the Trypsin depending on whether or not the cells are being maintained in FCS-free medium. Cells were centrifuged at 290 x g for 3 minutes (room temperature) and the pellet was resuspended in PBS. Cells were counted using a haemocytometer. 1×10^5 cells grown in each condition were stained with each antibody. The excess cells were either discarded at this point, or after fixation. Cells were pelleted at 290 x g for 3 minutes (room temperature), resuspended in 0.25 % (w/v) PFA dissolved in PBS at a concentration of 5×10^5 cells ml^{-1} , and fixed for 1 hour at room temperature with constant, gentle agitation. Cells were pelleted for 3 minutes at 450 x g (room temperature) and resuspended in 70% (v/v) ice-cold methanol in PBS. After 1 hour at 4°C with constant gentle agitation, cells were pelleted for 3 minutes at 450 x g (room temperature) and resuspended in 5% (v/v) serum. The choice of serum depends on the species used to raise the secondary antibody. Cells were washed once and resuspended in 5% serum at a final concentration of 1×10^6 cells ml^{-1} . 1×10^5 cells were used for each staining reaction. 100 μl aliquots of cells were incubated in blocking buffer for 30 minutes at room temperature and collected by centrifugation for 3 minutes at 450 x g at

room temperature. The cells were resuspended in 100 µl staining buffer containing the appropriate antibody and incubated at 4°C for 2 hours. Cells were pelleted (3 minutes at 450 x g at 4°C) and washed in pre-cooled staining buffer for 20 minutes at 4°C. The wash step was repeated twice. For primary conjugated antibodies, cells were resuspended in 200 µl PBS containing 5% serum and analysed by flow cytometry. Otherwise, cells were resuspended in 100 µl staining buffer with the appropriate secondary antibody and incubated for 2 hours at room temperature protected from the dark. Cells were pelleted by centrifugation (3 minutes at 450 x g at 4°C) and washed with 1 ml pre-cooled staining buffer for 20 minutes at 4°C. The wash step was repeated twice before cells were resuspended in 200 µl PBS containing 5% serum and passed through a 0.45 µm filter into a 12 x 75 mm polypropylene tube. Samples were submitted to the MRC flow cytometry facility for analysis using an Aria IIu (BD Bioscience) or Aria II (BD Bioscience) flow cytometer.

2.8 Mass spectrometry

2.8.1 DNA extraction

Genomic DNA for LC-MS analysis was isolated using the ZR-Duet DNA/RNA Miniprep kit (Zymo Research) according to the manufacturer's protocol, unless otherwise stated. Of note, the IIIC DNA binding columns were washed 3 times with 700 µl DNA/RNA wash buffer followed by a dry spin at 12,000 x g for 2 minutes before eluting the DNA in 50 µl LC-MS grade H₂O. DNA was quantified using a NanoDrop spectrophotometer and A₂₆₀/A₂₈₀ and A₂₆₀/A₂₆₀ ratio used as a measure of purity. DNA extracts with A₂₆₀/A₂₃₀ ratios below 1.5 were further purified using the DNA Clean and Concentrator kit (Zymo Research). DNA was eluted from the Zymo Spin Column using 15 µl LC-MS grade H₂O preheated to 55°C.

2.8.2 Purification of digestion enzymes

Custom-made DNA digestion enzyme (produced by NEB), was purified using Microcon 10 kDa centrifugal filters with Ultracell membrane (Merck Millipore). Briefly, to remove trace amounts of glycerine the filters were washed twice with 100 µl LC-MS grade H₂O at 14,000 x g until the membrane was *almost* dry (approximately 20 minutes). The filters were then washed with 50 µl 1X digestion buffer at 14,000 x g until the membrane was *almost* dry. Sufficient enzyme to digest all samples (0.2 µl per sample) was diluted in 50 µl digestion buffer, added to the centrifugal filter and centrifuged at 14,000 x g until the membrane was *almost* dry. The membrane was washed once more using 100 µl 1X digestion buffer. The purified enzyme was retrieved from the membrane by inverting the assembly, placing it into a clean collection tube and centrifuging to 1 minute at 1000 x g. The purified enzyme was diluted in 1X digestion buffer so that the total volume for each digestion reaction is 10 µl. Purified enzymes were used immediately for DNA digestion

2.8.3 DNA digestion and sample preparation

For small numbers of cells, (e.g. oocytes, zygotes and PGCs) all extracted DNA was digested to nucleosides. Otherwise, 250 ng DNA was digested per sample. DNA was initially vacuum concentrated in a DNA LoBind tube until < 1 µl remained. The concentrated DNA was subsequently resuspended in 10 µl 1X purified digestion enzyme mix and incubated at 37°C for 8 hours. The resulting nucleosides were diluted in LC-MS grade H₂O so that the amount of dG quantified was approximately 200 fmol.

2.8.4 Standard curve preparation

A dilution series of synthetic nucleoside standards was made according to Table 2.5. Samples and nucleoside standards were spiked with equal amounts of isotope labelled nucleosides, provided by Dr. T. Carell (LMU, Germany).

Table 2.5: preparation of LC-MS standard curve

Stock master mix 1 nmol/ μ l dC, 1 nmol/ μ l dG, 50 pmol/ μ l 5mdC, 50 pmol/ μ l 5hmdC, 50 pmol/ μ l 5fC				
Level	Concentration dC & dG	Concentration 5mdC, 5hmdC and 5fC	μ l Nucleosides	μ l LC-MS grade H ₂ O
1:100	10 pmol/ μ l	500 fmol/ μ l	2	198
Level 10	500 fmol/ μ l	25 fmol/ μ l	10	190
Level 9	100 fmol/ μ l	5 fmol/ μ l	100	400
Level 8	50 fmol/ μ l	2.5 fmol/ μ l	250	250
Level 7	10 fmol/ μ l	500 fmol/ μ l	100	400
Level 6	5 fmol/ μ l	250 fmol/ μ l	250	250
Level 5	1 fmol/ μ l	50 fmol/ μ l	100	400
Level 4	500 amol/ μ l	25 fmol/ μ l	250	250
Level 3	100 amol/ μ l	5 amol/ μ l	100	400
Level 2	50 amol/ μ l	2.5 amol/ μ l	250	250
Level 1	10 amol/ μ l	0.5 amol/ μ l	100	400

2.8.5 Nucleoside quantification

Samples were resolved on a Agilent RRHD Eclipse Plus C18 2.1 x 100 mm 1.8 μ column using an Agilent UHPLC 1290 system and detected using an Agilent 6490 triple quadrupole mass spectrometer. Agilent MassHunter Quantitative software was used to quantify the response and concentration of each sample based on the response of the known standards. To calibrate for the effects of ion suppression, standard curves representing the ratio of isotope labeled to unlabelled nucleosides against the ‘expected’ concentration were generated. This produced a calibration factor which could be applied to the peak-area values (response) of each sample.

2.9 Transcript analysis

2.9.1 Isolation of RNA.

When samples and cell numbers were not limiting, RNA was isolated using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Otherwise, RNA was isolated using the ZR-Duet DNA/RNA Miniprep kit (Zymo) according to the manufacturer's instructions. RNA was quantified using a Nanodrop spectrophotometer and the the A260/280 and A260/230 ratios recorded as a measure of their purity.

2.9.2 cDNA synthesis

Reverse transcription was performed using the PrimeScript RT reagent kit with gDNA eraser (Takara) according to the manufacturer's instructions. Minus reverse transcriptase controls for amplification from non-RNA templates were included. cDNA was diluted 10x before analysis using real-time quantitative PCR (RT-qPCR).

2.9.3 Real-time quantitative qPCR

To design RT-qPCR primers (Table 2.6), reference sequences were obtained from the UCSC genome browser. Primers were designed using Primer3 software so that the amplification spanned an intron. Off target amplifications were screened using UCSC in silico PCR tools. The amplification efficiency for each primer pair was determined using sequential 1:2:5 dilutions of cDNA. Primers with good linearity were selected for future RT-qPCR analysis.

Table 2.6: RT-qPCR primer sequences

Transcript	Forwards primer	Reverse primer
Dnmt1	CGGAACTAGGTGAAGTTTCAA	CAAATAGATCCCCAAGATCCA
Dnmt3a1	TTCTCCTCCTGTTCTCTCCT	GCCTACTGCCCAGCAATG
Dnmt3a1&2	TGATGATTGACGCCAAAGAA	TCACAGTGGATGCCAAAGG
Dnmt3b	GATTCTGGGGGAGGCTCTT	CAGGACGCGCTCTGAAAG
Dnmt3b total	GCGCAGCGATCGGCGCCGGAGAT	CATACCCGGTGGCACCCTCTTCTTCAGTCA
Dnmt3b catalytic	AAGCCCATGCAATGATCTCTCTAAC	CACGTCCGTGTAGTGAGCAGGGAAGC
Dnmt3b exons 10 and 11	CTGGAGAGTCACTGGAGGACCAGCTGA AGC	CTCTCCTCATCCTCCCCTCGGTCCTTC
Dnmt3L	AACATTCCAGGGCTGAAGAG	GCTTGCTCCTGCTTCTGACT
Tet 1	CCATTCTCACAAGGACATTCACA	GCAGGACGTGGAGTTGTTCA
Tet 2	GCCATTCTCAGGAGTCACTGC	CTTCTCGATTGTCTTCTCTATTGAGG
Prdm14	ACAGCCAAGCAATTTGCACTAC	TTACCTGGCATTTCATTGCTC
Oct4	GATGCTGTGAGCCAAGGCAAG	GGCTCCTGATCAACAGCATCAC
Sox2	CATGAGAGCAAGTACTGGCAAG	CCAACGATATCAACCTGCATGG
Nanog	CTTTCACCTATTAAGGTGCTTGC	TGGCATCGGTTTCATCATGGTAC
Uhrf1	GTTACAGCCTGTACTGAGGAAG	TAGAACTGTGCTGTCCAGTCTG
Smad1	CTGTCCCTCCTGTTCCAGC	GCCCGGAGCTAAGCAGAG
Smad4	GGCTGTCCTTCAAAGTCGTG	GGTTGTCTCACCTGGAATTGA
Smad5	CTCATAGGCGACAGGCTGA	AGATGGCCCCAGATAATTCC
Stella	AGGCTCGAAGGAAATGAGTTTG	TCCTAATTCTTCCCGATTTTCG
Pax6	TGAATGGGCGGAGTTATGAT	GGACGGGAACTGACACTC
T	GATGATTGTGACCAAGAACGG	CGAAGTCCAGCAAGAAAGAG
Gata4	GCCCAAGAACCTGAATAAATC	TAGTGGCATTGCTGGAGTTA
Sox7	AAGGATGAGAGGAAACGTCTG	ATCCACATAGGGTCTCTTCTG
Lefty 1	CCAACCGCACTGCCCTTAT	CGCGAAACGAACCAACTTGT
G6pdh	GCCACATGAATGCCCTGCAC	CGACAGTTGATTGGAGCTCTG

The RT-qPCR reaction mix consists of SensiMix SYBR Green No ROX (2X) qPCR master mix (Bioline), 0.3 mM RT-qPCR verified primers and 3 µl cDNA in a final reaction volume of 13 µl. RT-qPCR was performed using a CFX96 Real-Time qPCR machine (BioRad) using the following programme; activated by incubation at 95°C for 10 minutes followed by a cycle of denaturation at 95°C for 15 seconds, annealing at 60°C for 15 seconds and extension at 72°C for 20 seconds. This cycle was repeated 39 times; each time the SYBR fluorescence being read at the end of the extension step. A melting curve from 55°C to 95°C was performed at 1°C increments. The relative abundance of cDNA was calculated using the standard curve method. Briefly, cDNA diluted 1:2:5 was used as input for RT-qPCR. The average C(t) of three replicates was calculated for each level and used to generate a standard curve of input vs C(t) value. The input of the samples are then extrapolated from this curve. All values are normalised to a reference. The housekeeping gene *G6pdh* was chosen since its expression had been shown to be constant in mESCs maintained in FCS and 2i culture conditions.

2.9.4 Extraction of miRNA from mESCs

RNA enriched for miRNA was isolated from snap frozen mESC pellets using the mirVana miRNA Isolation Kit (Thermo) according to the manufacturer's protocol. Briefly, cell pellets were thawed on ice. 600 µl Lysis/Binding buffer was added to each sample. A homogeneous cell lysate was created by vortexing. The cell lysate was incubated with 60 µl miRNA Homogenate additive on ice for 10 minutes. 600 µl Acid-Phenol:Chloroform was added to the homogenate and mixed thoroughly by vortexing for 60 seconds. Samples were centrifuged at 10,000 x g at room temperature for 10 minutes to separate the aqueous and organic phases. The aqueous upper phase was transferred to a fresh tube without disturbing the organic phase. 1/3 volume of 100% ethanol was added to each sample and mixed thoroughly by pipetting. The lysate/ethanol mixture was applied to a mirVana filter cartridge. Isolated miRNAs were extracted from bulk miRNA by centrifugation at 10,000 x g for 15 seconds. 2/3 volume 100% ethanol was added to the filtrate and mixed thoroughly. The filtrate was then transferred to a second filter cartridge and centrifuged for 15 seconds at 10,000 x g. The filter cartridge was washed once with 700 µl mirVana wash solution 1 and twice with 500 µl wash solution 2/3. RNA was eluted from the filter cartridge using nuclease-free water pre-heated to 95°C into an RNase-free tube.

2.9.5 cDNA synthesis from miRNA enriched RNA

cDNA was synthesized from miRNA enriched RNA using the TaqMan Advanced miRNA cDNA synthesis kit according to the manufacturer's protocol. Briefly, 100 ng RNA was used for each poly(A) tailing reaction. A reaction cocktail consisting of 0.3 µl 10X poly(A) buffer, 0.5 µl ATP and 0.3 µl poly(A) enzyme in a total volume of 3 µl was incubated with 2 µl RNA at 37°C for 45 minutes. The poly(A) enzyme was heat inactivated at 65°C for 10 minutes. A ligation cocktail consisting of 3 µl 5X DNA ligase buffer, 4.5 µl 50% PEG 8000, 0.6 µl 25X ligation adaptor and 1.5 µl RNA ligase in a total volume of 10 µl was thoroughly mixed with the poly(A) tailing reaction and incubated at 16°C for 1 hour. The ligation reaction mixture was immediately mixed with a reverse transcription cocktail consisting of 6 µl 5X RT buffer, 1.2 µl dNTP mix (25 mM each), 1.5 µl 20X universal RT primer, 3 µl 10X RT enzyme and 3.3 µl RNase free dH₂O. The samples were incubated in a thermocycler for 15 minutes at 42°C. The reverse transcriptase was heat inactivated at 85°C for 5 minutes. Universal forwards and reverse primers were used to increase the total number of cDNA molecules as follows; a miR-amplification cocktail consisting of 25 µl 2X miR-Amp master mix and 2.5 µl 20X miR-Amp primer mix in a total volume of 45 µl dH₂O was added to 5 µl of the reverse transcription reaction product and thoroughly mixed. The polymerase was heat activated by incubation at 95°C for 5 minutes followed by 15 cycles of denaturation at 95°C for 3 seconds and annealing/extension at 60°C for 30 seconds. The polymerase was heat inactivated by incubation at 99°C for 10 minutes.

2.9.6 miRNA RT-qPCR

cDNA was diluted 1:10. For each reaction 5 µl cDNA was added to 15 µl RT-qPCR cocktail consisting of 10 µl 2X Fast advanced master mix, 1 µl 20X Taqman Advanced miRNA assay and 4 µl RNase free dH₂O.

2.10 Protein expression

2.10.1 Protein extraction

Cells were either directly lysed in ice-cold RIPA buffer (150 mM sodium chloride, 1.0% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecylsulfate, 50 mM Tris pH 8.0) or they were pelleted and snap frozen in liquid nitrogen. Protease (and phosphatase) inhibitors were added to RIPA buffer prior to use. After resuspension in RIPA buffer, samples were placed on ice and constantly agitated for 30 minutes. To increase protein yields, the samples were sonicated for 5 cycles (30 seconds on, 30 seconds off) on high power. The cell lysate were transferred to a clean 1.5 ml tube and centrifuge at 4°C 20,000 x g for 15 minutes to remove any cell debris.

2.10.2 Protein quantification

Protein was quantified using a bicinchoninic acid (BCA) protein assay (Thermo 23227). Briefly, a serial dilution of BSA from 2000 µg/ml to 25 µg/ml was prepared. Standards were run in triplicate along with three blanks. A working reagent was prepared according to the manufacturer's instructions (using BCA reagent A with BCA reagent B in a 50:1 ratio). 200 µl working reagent was used for each sample. 25 µl of each standard or sample was added in replicate to a 96 well flat bottomed plate, followed by 200 µl of the working reagent. The plate was incubated for 30 minutes at 37°C. The absorbance was read at 562 nm after allowing the plate to cool to room temperature.

2.10.3 SDS PAGE and Western Blot

6x Laemmli buffer (12% SDS, 60% glycerol, 375 mM Tris-HCl (pH 6.8) and 0.015% bromophenol blue) was added to 20 µg of each protein sample. Protein extracts were denatured for 5 minutes at 99 °C and then chilled on ice. The protein extracts were loaded onto a SDS polyacrylamide gel consisting of a 4% stacking gel (4% (w/v) acrylamide, 125 mM Tris-HCl (pH 6.8), 0.1% SDS, 0.067% ammonium persulphate (APS), 0.12%

N,N,N'-tetramethylethylenediamine (TEMED)) and a 10% acrylamide resolving gel (10% (w/v) acrylamide, 390 mM Tris-HCl (pH 8.8), 0.1% SDS, 0.05% ammonium persulphate (APS), 0.1% N,N,N'-tetramethylethylenediamine (TEMED)). A protein ladder was run alongside the protein extracts as a size reference (Benchmark prestained protein ladder). Proteins were separated in running buffer (25 mM Tris base, 192 mM glycine, 0.1% SDS) using a Bio-Rad minigel system for 1 - 1.5 hours at 150 V. The proteins were transferred to a Hybond-P polyvinylidene fluoride transfer membrane (Amersham Biosciences) using a mini Trans-Blot electrophoretic transfer cell (Bio-Rad) in transfer buffer (48 mM Trizma base, 39 mM glycine, 0.037% (w/v) SDS and 20% (v/v) methanol). Each transfer was performed at 250 mA for 1.5 hour. Membranes were blocked for 30 minutes in blocking buffer (5% (w/v) fat-free milk powder, 0.1% Tween 20 in TBS or 5% (w/v) BSA, 0.1% Tween 20 in TBS) followed by overnight incubation with the primary antibody in the same buffer at 4°C with constant agitation. The membranes were washed three times with wash buffer (0.1% Tween 20, TBS) for 10 minutes before a 1 hour incubation with a horseradish peroxidase-coupled secondary antibody diluted in blocking buffer at room temperature. After washing the membrane with three times with TBS for 10 minutes, the protein was detected using Luminata Crescendo Western HRP substrate (EMD Milipore).

Table 2.7: Primary antibodies used for Western blots

Antibody	Source	Dilution
anti-Dnmt1	Imgenex IMG-261	1:500
Anti- Dnmt3a	Imgenex IMG-268A	1:500
Anti-Dnmt3b	Imgenex IMG-184A	1:500
Anti-Prdm14	Millipore AB4350	1:500
Anti-RAR α	Santa Cruz sc-551	1:200
Anti-RAR β	Santa Cruz sc-773	1:200
Anti-RAR γ	Santa Cruz sc-7387	1:200
Anti-CRABP II	Millipore MAB5488	1:500
Anti-phospho Akt T308	Cell signalling 2956	1:1000
Anti-phospho Akt S473	Cell signalling 4060	1:1000
Anti-lamin B	Santa Cruz sc-6216	1:10,000

2.11 Bacterial protocols

2.11.1 Transformation of bacterial cells

One Shot TOP10 Chemically Competent *Escherichia coli* (*E. coli*) (Invitrogen) were incubated with 10-100 ng plasmid (no greater than 5% bacterial volume), on ice for 30 minutes. The transformation mixture was heat-shocked at 42°C for 30 seconds on a heat block without shaking and cooled rapidly on ice for a further 2 minutes. 250 μ l prewarmed S.O.C rescue media was added and the transformation mixture was incubated at 37°C with shaking at 225 rpm for 1 hour. 2-50 μ l transformed cells were aseptically spread onto L-agar plates with the appropriate antibiotic added (Ampicillin at 50 mg/ml, Kanamycin at 50 mg/ml) using sterile glass beads and incubated at 37°C overnight.

2.11.2 Bacterial culture and isolation of plasmid DNA

For the preparation of plasmid DNA, transformed TOP10 cells were incubated at 37°C overnight with shaking (225 rpm) in 2 ml liquid Luria-Bertani (LB) culture medium containing the appropriate antibiotic. For maxi-preparations, 1 ml of this starter culture was inoculated into 150 ml of antibiotic containing LB and incubated at 37°C for a further 16 hours. Plasmid DNA was extracted with either the Qiagen Plasmid Maxi kit (Qiagen) or the Qiagen Plasmid Mini kit (Qiagen) according to the manufacturer's instructions.

Chapter 3

Quantification of global DNA modifications

by LC-MS

3.1 Introduction

Various approaches have been developed to quantify DNA methylation levels. These techniques can broadly be divided into those which quantify 5mdC at the global or sequence specific level (Reviewed in Fraga & Esteller, 2002) (Plongthongkum et al., 2014). Techniques which are sequence specific can further be divided into those looking at the entire genome wide level and those that focus on a specific locus. Regarding the approaches that measure global DNA methylation, high performance liquid chromatography (HPLC) was first used to quantify 5mdC in 1980 (Kuo et al., 1980). Although robust, this method relies on the availability of large quantities of DNA (3-10 μ g); thus being wholly unsuitable for the study of embryonic development where technical and ethical considerations limit the amount of biological material. Technical advances have since led to the coupling of HPLC with more sensitive and selective detection methods, including mass spectrometry. Liquid chromatography-mass spectrometry (LC-MS) today is an invaluable tool for the analysis of biomolecules including peptides, metabolites and nucleotides.

3.1.1 Reverse-phase chromatography

LC-MS analysis comprises a multi-step workflow. Developing a highly sensitive LC-MS method relies on the optimisation of each stage. For example, resolving the analytes by HPLC, reduces the interaction between different molecular species during the downstream ionisation process, aids the identification of biomolecules based on retention time and facilitates the optimisation of parameters, including Δ EMV, to increase the sensitivity of detection for each individual species. Typically, this first step of the HPLC workflow relies on separation of analytes based on hydrophobicity (reverse-phase chromatography). Broadly speaking, hydrophobic analytes are adsorbed from a polar mobile phase onto a hydrophobic stationary phase. The adsorbed hydrophobic molecules can be subsequently eluted using a gradient of organic solvent. Chromatographic separation aims to resolve narrow peaks with high signal-to-noise ratio (SNR) in a short time

frame and with sufficient sensitivity to detect all analytes of interest. A generic reverse-phase HPLC column consists of an electro-polished stainless steel tube filled with silica particles.

Choosing the correct stationary phase (column) can be a simple means of improving peak shape and resolution. Simplistically, the reverse-phase ligand determines the mechanism by which a solute is retained on the stationary phase and thus the resolution of solute peaks. For instance, unmodified alkyl ligands (including C18, C8 and C4) bind to analytes via polar dispersive Van der Waal interactions such that retention is proportional to analyte hydrophobicity whereas phenyl ligands interact primarily with an analyte's polar functional groups. As such, the choice of stationary phase should be guided by the chemical differences between the analytes of interest.

The choice of column is complicated by the nature of the silica itself. Traditionally, column supports have consisted of complex porous silica microspheres, for which the interaction between the analyte and stationary phase includes the diffusion of the analyte into the porous packing followed by multiple rounds of retention and elution from the silica. Alternatives to porous microparticles, including perfusion particles and micropellicular particles, have been developed to increase the efficiency of separation and to reduce the width of chromatographic peaks. Particle shape and size are also important variables; irregular shaped silica particles provide a greater surface area for stationary phase-analyte interactions whilst spherical particles tend to yield greater reproducibility. Reproducibility is also negatively affected by broad particle size distributions (Reviewed in Wu et al., 2007).

Mobile phases used in reverse-phase chromatography are polar to ensure that hydrophobic analytes are initially retained on the hydrophobic stationary phase (column). The proportion of aqueous to organic solvent is then typically reduced over the course of a single run such that species are individually eluted from the stationary phase of the column. Therefore, HPLC method development necessitates the running a 'scouting' gradient to gauge the hydrophobicity of the mobile phase required to elute each of the analytes of interest. Alternatively,

isocratic elutions, in which the concentration of aqueous to organic solvent does not change, can lead to greater resolution between structurally similar analytes.

3.1.2 Electrospray ionisation

The compounds eluted from the HPLC column are then transmitted into the mass spectrometer using atmospheric pressure ionisation methods including electrospray ionisation (ESI) or atmospheric pressure chemical ionisation (APCI) (Fenn et al., 1989). The experiments described in this thesis exclusively utilise an ESI source. Transferring ions from a solution to a gas phase crucially depends on three steps; the production of charged droplets, Coulombic fission and finally the formation of gaseous ions. In the positive ion mode the capillary forms the anode whilst the sampling aperture becomes the cathode. Positively charged ions are repelled from the walls of the nebuliser needle towards the capillary tip (Fig. 3.1). The resultant increase in charge density at the tip of the capillary increases the repulsive (Coulombic) forces. When the Coulombic forces are greater than the surface tension of the eluent, the formation of a Taylor cone in place of the meniscus reduces these repulsive forces. Electrospray occurs from the tip of the Taylor cone as droplets with a vast excess of anions are expelled from the eluent. This process is known as nebulisation. Expelled droplets enter the desolvation zone; a region of high ambient air temperature, often aided by the use of a heated, inert desolvation 'drying' gas such as pressurised nitrogen. Solvent evaporation reduced the droplet radius. Since the overall charge does not change, solvent evaporation reduces the inter-ionic distances, thus increasing repulsion between ions in a single droplet. When the droplet surface tension (Rayleigh instability limit) is exceeded, jet fission will reduce the Coulombic stress. The resulting droplets rapidly undergo further desolvation-fission cycles ultimately producing droplets of approximately 10 nm radius, each with a small number of charges.

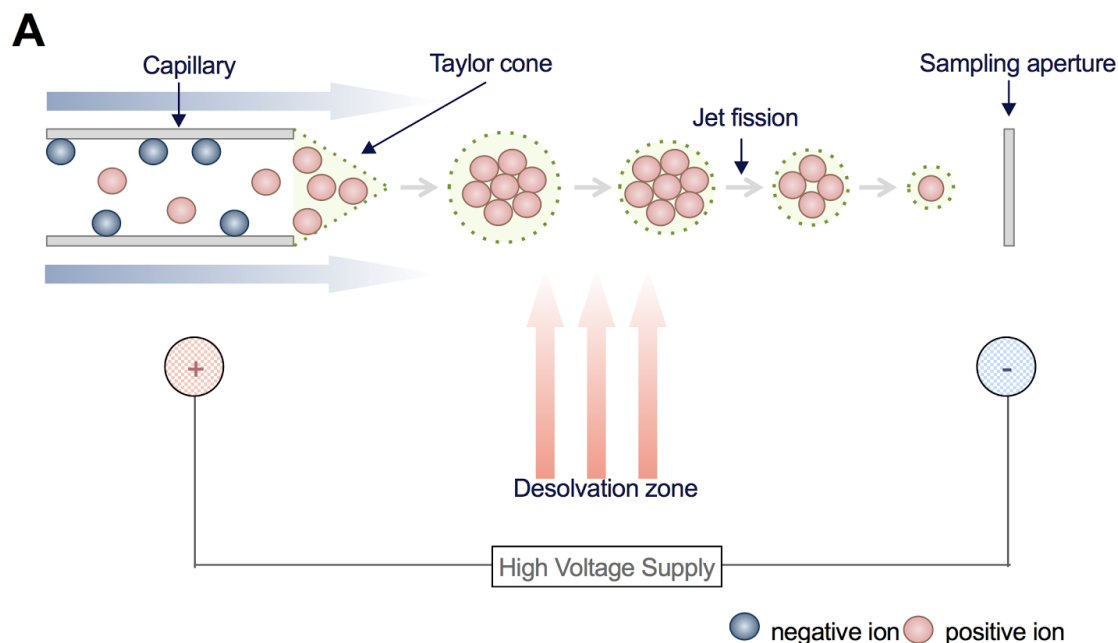


Figure 3.1 Electrospray ionisation

(a) Analytes eluted from the chromatographic column enter the electrospray chamber via a nebuliser needle. The mode of the mass spectrometer (positive or negative) determines whether the charged droplets contain negative or positive ions. Nebulisation occurs from the tip of the Taylor cone. Repeated solvent evaporation and jet fission reduce the droplet size as they pass through the desolvation zone *en route* to the sampling aperture.

Gas-phase ions enter the mass spectrometer via the sampling aperture. In the Agilent 6490 triple quadrupole system used throughout these experiments, 6-times the number of ions of conventional single capillary setups are captured; therefore enabling a much greater proportion of ions at the center of the nebulised plume to be sampled. Subsequently these ions are captured by the dual ion funnels, which reduce neutral noise and significantly improve signal-to-noise by separating ions from atmospheric gas .

3.1.3 Triple quadrupole mass spectrometry

Quantification of the ionic species depends on their separation according to their mass-to-charge ratio. This is most commonly achieved using quadrupoles. The first quadrupole was described and constructed by Wolfgang Paul and Helmut Steinwedel. The basic principle relies on the separation of ions as they pass along the central axis of two pairs of parallel rods. The two rods lying along the x-axis are diametrically opposed to the two rods

lying along the y-axis. These rods act as pairs of electrodes to which a direct current (DC) and alternating current (AC) are applied. By altering the magnitude of the applied electric field, ions with a particular m/z can be selected to pass through the quadrupole on a stable trajectory whilst those on unstable or collisional trajectories are lost.

Increasingly, basic mass spectrometry is being replaced by tandem mass spectrometry (MS/MS). This simply means combining two or more MS experiments (either in time or space) and is typically employed to either provide structural information by fragmenting the ions from the first MS, or to increase selectivity and sensitivity by specificity by detecting only the species with specific m/z ratios in MS1 and MS2. The Agilent 6490 is a triple quadrupole (QQQ) consisting of two quadrupoles separated by a collision cell. Triple quadrupoles routinely enable three scanning modes (Fig. 3.2); multiple reaction monitoring, product ion scan and precursor ion scan. The first of these is typically used to quantify analytes of known m/z and fragmentation properties whilst the latter are useful during method development and identification of novel molecules.

3.1.4 Ion suppression, sensitivity and reliability

The importance of sample preparation is often overlooked when developing LC-MS methods. Sample cleanup removes potential interferents by separating the compounds of interest from a complex sample matrix at a concentration that is sufficient for detection and quantification. Additionally, insufficiently pure samples can lead to reduced experimental reproducibility and have adverse effects on the useful life of the column, therefore increasing the need for frequent cleaning of the injector and detector modules.

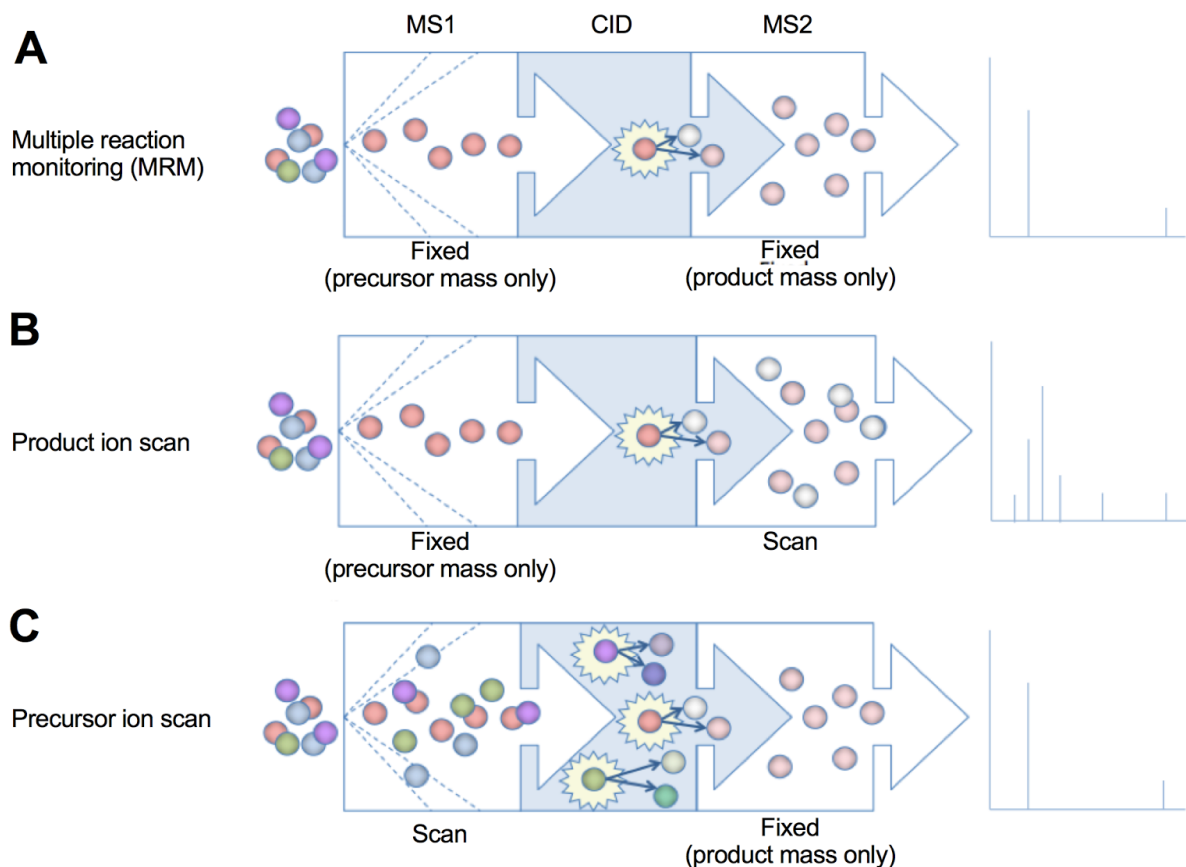


Figure 3.2 Schematic of the common scanning modes used during QQQ mass spectrometry

(a) Multiple reaction monitoring scans are used when the mass of the precursor and product ion are known. Ions are filtered based on their m/z ratio in both Q1 and Q3. Analysing only a small number of transitions increases the sensitivity of detection. (b) Product ion scans select a precursor ion based on its known m/z and scans for all product ions produced following collision induced dissociation (CID). This mode is useful during the optimisation of the collision energy. (c) Precursor ion scans can identify all ions in MS1 which fragment to produce a particular product ion of known m/z .

Additionally, contaminants within the sample preparation are widely responsible for matrix effects; effects caused by components other than the analyte of interest. For instance, ion suppression is caused by the coelution of solutes which interfere with the ionisation process (Kearle & Tang et al., 1993)(Taylor, 2003)(Matuszewski et al., 2003)(Annesley, 2003)(Larger et al., 2005). There are numerous sources of ion suppression, including compounds endogenous to the sample and those which have been introduced through the sample preparation process. Typically, this phenomenon results in a loss of signal response; although it is

also possible that the observed response may be overestimated (Matuszewski et al., 2003). Ion suppression is most often a concern when small amounts of analytes are being extracted from a complex biological sample.

Although the precise mechanism of ion suppression is unknown, the responsible contaminants are thought to interfere with nebulisation and jet fission. For example, the Rayleigh instability limit can be increased by increasing the surface tension of the droplets, thereby reducing solvent evaporation (desolvation) and the ability of ions to enter the gas phase. Thus, any compound that increases the viscosity of the droplets is a source of ion suppression. Alternatively, non-volatile substances can reduce ion expulsion by preventing droplets reaching their critical radius.

Various strategies can be employed to reduce the effects of ion suppression, including either diluting the sample or reducing the injection volume. This approach, however, is clearly inappropriate for samples with small amounts of biological material or for identifying modifications of low abundance. Alternatively, optimising the sample preparation may eliminate the culprit contaminants. Since the source of ion suppression can be endogenous or exogenous (i.e., originating from the digestion enzymes, buffers etc), it is important to consider purification of all substances which could possibly enter the ESI chamber, for instance components of the DNA extract or the digestion enzymes. Increasing the selectivity of the chromatographic separation may also reduce any observed matrix effect by resolving analytes from potential sources of ion suppression. Alternatively, isocratic elutions may resolve overlapping contaminating peaks and therefore could reduce ion suppression. Finally, when matrix effects cannot be eliminated, steps can be taken to account for their impact on the analyte response. Stable isotope labelled synthetic nucleoside standards will have identical retention times and ionisation properties to the sample nucleosides. Therefore, by adding a known concentration of isotope labelled synthetic nucleosides to the calibration curve and samples, we can calculate the effect of ion suppression and subsequently correct the measured response to generate an accurate measurement (Duxbury et al., 2008)(Stokvis et al., 2005).

3.2 Experimental aims

I wanted to further optimise our LC-MS method to facilitate highly sensitive quantification of DNA modifications in small cell numbers. Specifically, we believed that ion suppression significantly limits the accuracy, reliability and robustness of our current methods. Therefore, the development and employment of additional protocols to reduce or correct for matrix effects are required. Specifically, we wanted to answer the following questions:

- What are the source(s) of ion suppression?
- Does eliminating these interferents improve the accuracy and reliability of our analysis?
- Can these methods be applied to the analysis of DNA modifications in small numbers of cells?

The utility of LC-MS can be greatly enhanced when combined with upstream applications, including flow cytometry to isolate specific cell populations. Typically, fluorescent activated cell sorting (FACS) relies on the fixation and permeabilisation of cells to enable the staining of intracellular antigens or, in the case of EdU labelling, the covalent addition of a fluorophore to a nucleoside analogue. Fixatives alter the molecular makeup of cells in numerous ways and therefore can have remarkably different effects on downstream applications. Any fixation steps used prior to DNA extraction must be compatible with our LC-MS methods. Therefore, I also wanted to assess the effects of different cytological fixatives on the downstream LC-MS analysis.

3.3 Results

3.3.1 Prologue

Previous members of our group have invested significant time developing a general LC-MS method for the detection and quantification of common nucleosides (dG, dC, 5mdC, 5hmdC, 5fC and 5CaC) in small number of cells, including PGCs, oocytes and zygotes (Leitch et al., 2013)(Nashan et al., 2015)(Amouroux et al., 2016). This process involved optimising peak resolution, the polarity, capillary (nebuliser) voltage, solvent system, flow rate, drying gas parameters, collision energy and the dwell time between measuring ion transitions. These parameters are summarised in Table 3.1 and Table 3.2. A combination of hardware design, acquisition optimisation and sample preparation steps have greatly improved the sensitivity of these methods. The steps taken to optimise several of these parameters are outlined below:

Table 3.1: Summary of flow dependent LC-MS parameters

Parameter	Details
Mode	Positive electrospray ionisation
Gas Temperature	130 degrees
Gas flow	20 L/min
Nebuliser	55 psi
Sheath gas temperature	400 degrees
Sheath gas flow	12 L/min
Capillary voltage	Positive 2000V, Negative 3000V

Table 3.2: Summary of compound-dependent parameters

Species	Precursor m/z	Product m/z	Collision energy	Dwell time	Cell accelerator voltage	Δ EMV
5fdC H	256	140	6	30	5	300+
5fdC Na	278	162	8	100	5	300+
5fdC K	294.1	178.1	8	15	5	300+
5hmdC H	258	142.1	6	30	2	500+
5hmdC Na	280.1	164.1	10	100	5	500+
5hmdC K	296.1	180.1	10	15	5	500+
5mdC H	242.1	126.1	6	30	5	400+
5mdC Na	264.1	148.1	10	100	5	400+
5mdC K	280.1	164.1	8	30	5	400+
dC H	228.1	112.1	6	30	6	500+
dC Na	150.1	134.1	8	80	5	500+
dC K	266.1	150.1	9	15	2	500+
dG H	268	152	6	30	5	300+
dG Na	290.1	174.1	10	80	5	300+
dG K	306.1	190.1	8	15	5	300+

I. Selection of a HPLC column based on nucleotide resolution and sensitivity

Synthetic nucleoside standards were used to determine the optimal elution conditions from 5 HPLC columns (4 reverse-phase and 1 normal phase) with unique chemical and physical properties (summarised in Table 3.3). Optimising peak resolution can be achieved using gradient elution, isocratic elution or a mixture of the two. MRM chromatograms illustrating the elution of the synthetic nucleosides clearly showed that the Zorbax Bonus RP and Zorbax SB Phenyl columns are unable to retain dC, 5mdC and hmdC; all three analytes elute from the column prior to the onset of the organic gradient. Zorbax Eclipse Plus C18 Rapid Resolution High Definition

(RRHD), HILIC and Hypercarb columns all retained the nucleosides until the onset of the gradient; therefore demonstrating that analyte retention can be controlled by solvent composition.

Table 3.3: Properties of HPLC columns tested

Column	Chromatography	Bonded (Stationary) Phase	Packing	Pore size	pH Range	End capped	Comments
Zorbax Bonus RP	Reverse-phase	Diisopropyl-C14	Uniform, spherical, ultra high purity Rx-SIL	80 Å	2.0-9.0	Triple	Designed to reduce interactions with polar compounds
Zorbax SB Phenyl	Reverse-phase	Diisopropyl-phenylsilane	StableBond porous microsphere	80 Å	1.0-8.0	No	Especially suited for use with high sensitivity detectors (e.g. MS) that require low background
Zorbax Eclipse Plus C18 RRHD	Reverse-phase	Dimethyl-n-octadecylsilane	Uniform, spherical, ultra high purity Rx-SIL	95 Å	2.0-9.0	Double	Superior peak shape with basic compounds
HILIC	Normal-phase	ZIC-HILIC Hydrophilic, Zwitterionic	Porous silica	200 Å	3.0-8.0	No	High affinity for polar compounds (e.g. nucleosides)
Hypercarb	Reverse-phase	Porous graphitic carbon	Spherical, porous, graphitic carbon	250 Å	0-14.0	No	Separates polar compounds with closely related structures

These three columns were then assessed to determine whether they were compatible with detecting low concentrations of analytes using MS. The sensitivity of MS is generally described in terms of limits of detection (LOD) and limits of quantification (LOQ). The LOD is the lowest analyte concentration that can be reliably distinguished from background noise. Ensuring that a LOD is meaningful requires a minimum signal-to-noise ratio (SNR) to be specified. For our analysis, noise is quantified using peak-to-peak measurements (minimum to maximum value of baseline noise). To standardise our LOD measurements, we have arbitrarily specified that the SNR must be at least 10.

Generally the LOD will lie somewhere below the assays functional sensitivity. In this case the LOQ, which represents the lowest concentration at which an analyte can be both reliably detected and accurately quantified, is a more meaningful value. LOQs are determined by comparing the measured response with the expected response. Typically a value which is +/- 20% from the expected response is considered to be an accurate measurement. The expected response is calculated based on the response measured for a serial dilution of nucleoside standards.

The sensitivity of 5mdC, 5hmdC and 5fC detection was much higher when using a Zorbax Eclipse Plus C18 RRHD column compared to either the HILIC or Hypercarb columns. It was suggested that nucleosides were retained on the HILIC and Hypercarb column, as evidenced by peak broadening and carry-over of analytes into the blank injections. Therefore, for all subsequent experiments, analytes were separated using the Zorbax Eclipse Plus C18 RRHD column.

II. Optimising solvent conditions

Our solvent conditions are based on a published protocol (Song et al., 2005). Developing solvent conditions for HPLC depends on selecting the correct organic solvent and mobile phase additives. Methanol and acetonitrile are by far the most common LC-MS solvents and differ in terms of polarity, viscosity, toxicity and cost. For instance, methanol is characterised as protic; meaning that it can create polar and ionic interactions with the analytes. For polar compounds such as nucleosides, this property forms the basis of greater selectivity, albeit at the expense of longer run times. Therefore, most of our methods use 100% H₂O and 80% methanol.

The mobile phase pH has a dramatic effect on the analyte ionisation state and therefore is one of the most important variables to optimise and control during LC-MS method development. Since nucleosides are basic, they will be protonated at a pH below their pK_a and neutral (or ion suppressed) above their pK_a . pH can be controlled by buffers (including ammonium formate and ammonium bicarbonate) or acids (typically formic acid). It is crucial that mobile phase additives used for LC-MS are volatile, and therefore many other common additives used for HPLC are unsuitable for use in our methods. While critical for the optimal MS analysis, the

ionisation state is inextricably linked to analyte retention on the HPLC column (with acidic or basic species being retained for shorter periods than their neutral counterparts). Therefore, the addition of formic acid is likely to reduce the retention time of nucleosides. However, when weighed against the substantially improved peak shape in the MS analysis, overall formic acid was thought to be an important addition to our mobile phase.

III. Optimisation of sample digestion

The sensitivity and reliability of LC-MS analysis is dependent on the quality of the sample. Therefore, developing a reliable means of sample preparation is an essential step in LC-MS method development. Specifically, sample preparation must ensure that the DNA digestion is complete with minimal loss of nucleosides, whilst also eliminating endogenous and exogenous interferents which negatively impact the analysis or have the potential to damage the triple quadrupole instrument.

In order to analyse the amount of modified DNA bases, the isolated DNA is digested to nucleosides. Before I joined the group, three commercially available enzymes (NP1, Benzonase and Degradase) were tested for their ability to digest DNA extracted from small numbers of cells (50, 100, 500 and 1000). Reliably quantifying DNA modifications from small cell numbers is technically challenging and all three enzymes displayed significant ion-suppression. The source of this ion suppression was initially attributed to the salts present in the enzyme buffer. However, dialysis of the digestion mix did not have a discernable effect on ion suppression. Additionally, spiking the samples with isotopically labelled nucleosides to increase the nucleoside:enzyme ratio did not improve the reliability.

3.3.2 Tackling ion suppression and enhancement effects

Given this, one of my first challenges was to optimise our sample preparation method to aid the reliable quantification of DNA modifications from small numbers of cells. As mentioned above, numerous different approaches can reduce the effects of ion suppression. The simplest of these is sample dilution. Figure 3.3b shows the effect of serial dilutions on the quantification of nucleosides extracted from E14 mESCs grown in conventional FCS conditions and digested using Degradase plus (Fig. 3.3a). In our analysis, ion suppression can

be identified as a deviation from the expected dC/dG ratio of 1.0. Although the dilution of biological samples does improve the dC/dG ratio, the dC response is still substantially lower than expected. In this particular scenario, ion suppression may be caused by factors present in the DNA extract (endogenous), or in the digestion matrix (exogenous). To determine whether the source of ion suppression was predominantly endogenous or exogenous, the LC-MS analysis was repeated using dilutions of Degradase Plus digestion matrix and synthetic nucleoside standards. Once again, dilution substantially improved the reliability of our method by reducing the effects of ion suppression. Notably, the effects of ion suppression are reduced when synthetic nucleosides rather than extracted DNA is analysed, indicating that ion suppression is caused by both endogenous (isolated genomic DNA) and exogenous sources (digestion buffer and enzyme). To date, I have used this approach to quantify DNA modifications in a wide variety of samples, including mouse cell lines, maize and *S. pombe* (Fig. 3.3c).

However, when quantifying DNA modifications in samples numbers of small cells, the concentration of genomic DNA is very low. Thus, further dilution of the nucleosides is not possible. As an alternative approach, endogenous and exogenous contaminants may be eliminated during the sample preparation. Typically, DNA is isolated by either organic extraction (phenol-chloroform) or one of many silica-based membrane technologies. These DNA isolation kits differ on several levels; including the lysis buffers required to separate the nucleic acids from the other cellular constituents, the composition of the denaturing solvents, the nature of the binding agent enabling the nucleic acid to interact with the silica (including the mixture and concentration of chaotropic salts) along with the salts, alcohol and chelating agents present in the wash buffer. The purity of the isolated DNA therefore depends not only on the ability to separate DNA from other cellular constituents, but also on the amount of carry over of these solvents and particulates. Whilst these differences may have only a minor impact on many downstream applications, the sensitivity of our LC-MS methods requires that each extraction protocol be assessed for its contribution to ion suppression.

To assess this, I isolated DNA from 1000 cells using either phenol-chloroform, the Qiagen DNeasy blood and tissue kit, or the Zymo DNA/RNA duet extraction kit with one of two silica columns (known as IIC or IC), digested the DNA using the Degradase Plus protocol and quantified the response of dC and dG as a measure of

ion suppression (Fig. 3.4). I found that ion suppression was greatest when samples were extracted using phenol-chloroform (Fig 3.4a) and, unsurprisingly, the response of all species was the lowest, suggesting a significant amount of biological material had been lost during the extraction process (Fig 3.4b). There was also substantial ion suppression seen in samples extracted by the Qiagen DNeasy blood and tissue kit. The Zymo Duet IIIC and IC columns are essentially the same, however, the amount of silica is reduced in the IC column. I found that ion suppression was the lowest in samples extracted using the IC columns, suggesting that components from the silica substrate can contribute to the observed ion suppression.

Interferents can also originate from the digestion mix. Glycerol, a common additive to enzyme storage buffers, is known to reduce ionisation during electrospray and is therefore an obvious source of ion suppression. To test the effects of glycerol on ion suppression, we have been given access to a series of custom-made enzymes, differing only in their glycerol concentration, produced by New England Biolabs. Using a standard mix comprising 100 fmol dG and dC and 25 fmol 5mdC and 5hmdC, I added 0.2 μ l of each enzyme in 1X digestion buffer in a total volume of 10 μ l. Additionally, I purified one of the enzymes using a centrifugal filter unit to eliminate any remaining glycerol. The effect of the digestion matrix on ionisation was analysed by comparing the dC/dG ratio between the samples (Fig. 3.5a). The use of Degradase Plus created by far the greatest suppressive effect, whilst ion suppression was lowered by the use of enzymes with reduced glycerol content. Therefore, for all subsequent studies, DNA was digested using glycerol-free NEB custom made enzymes, or the NEB custom made enzymes with the additional centrifugal purification step.

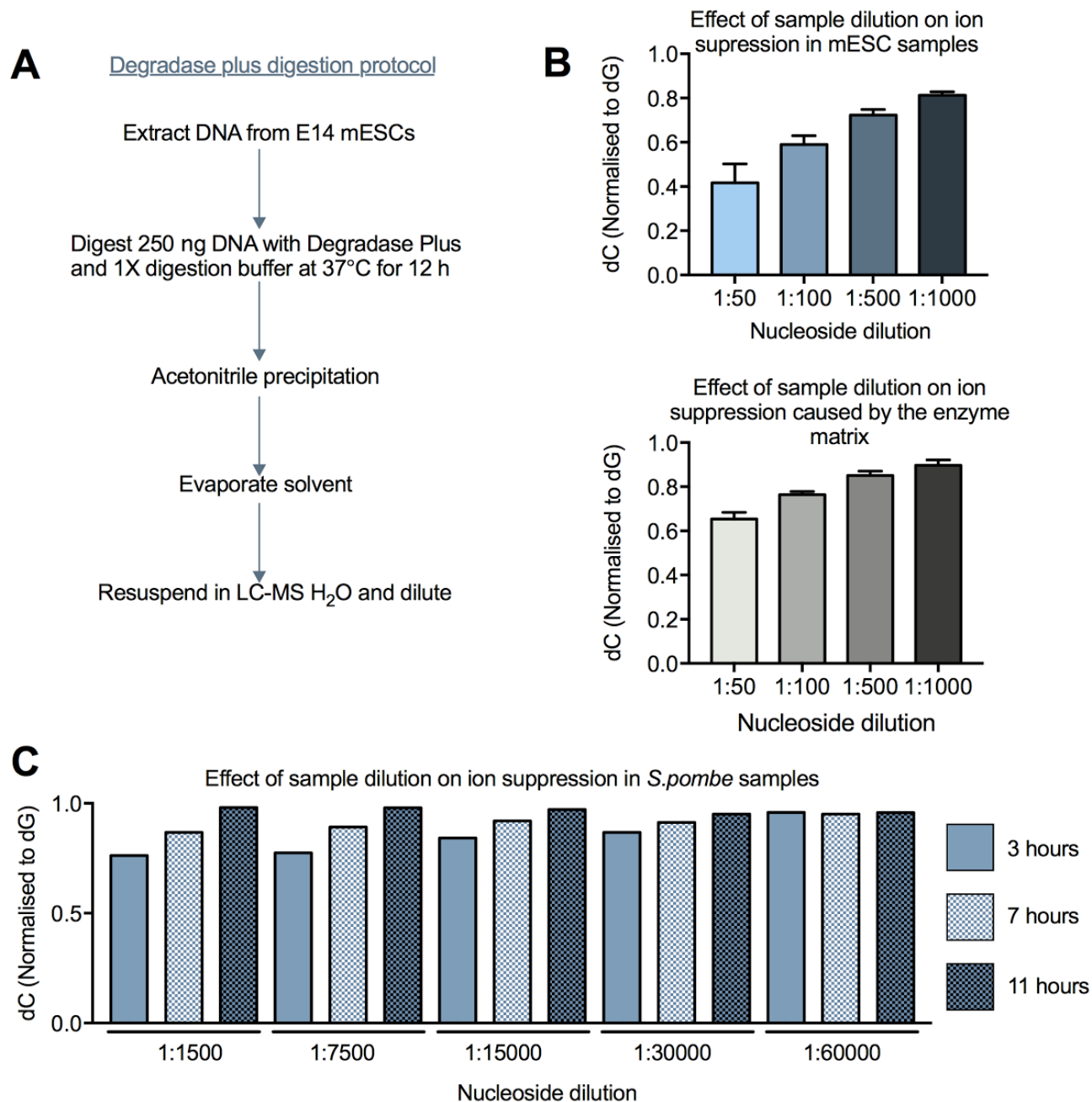


Figure 3.3 Sample dilution and the reduction of ion suppression

(a) Brief outline of the key steps of sample preparation using Degradase Plus enzyme mix.

(b) Sample dilution can effectively reduce ion suppression; measured as a deviation of the dC/dG ratio from 1. (Top) 250 ng DNA extracted from E14 mESCs was digested to nucleosides using Degradase Plus. Samples were diluted 1:50, 1:100, 1:500 and 1:1000 prior to LC-MS analysis. All measured responses were within the linear range. (Bottom) Degradase Plus digestion mix was added to dG and dC (5 pmol each). Samples were diluted 1:50, 1:100, 1:500 and 1:1000 prior to LC-MS analysis. All responses were within the linear range. In each instance, three biological replicates were analysed in duplicate.

(c) Dilution of the nucleoside preparation was used to reduce the effects of ion suppression during the quantification of DNA modifications that occur during the course of zygotic meiosis in *S. pombe*. Samples collected 3 hours, 7 hours and 11 hours after the onset of *S. pombe* meiosis were digested and diluted 1:1500, 1:7500, 1:15000, 1:30000 and 1:60000 before LC-MS analysis.

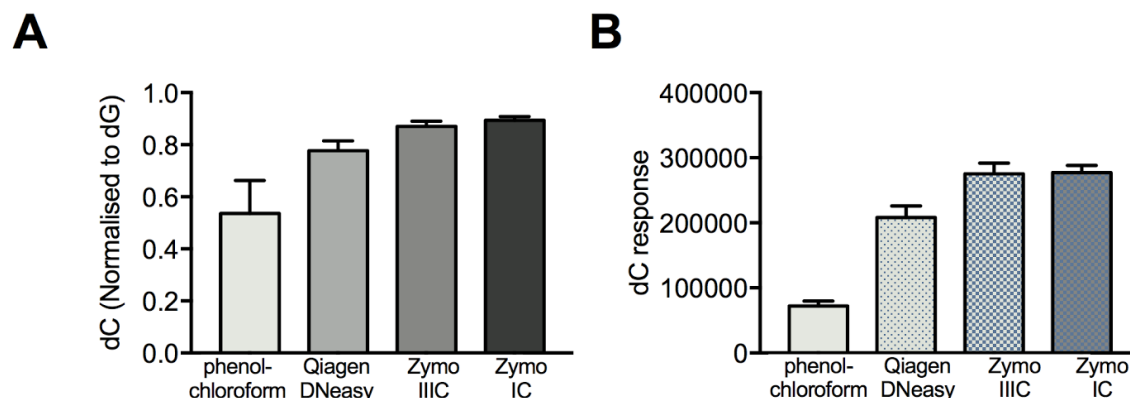


Figure 3.4 Ion suppression and the effect of DNA extraction methods

(a) DNA was extracted from 1000 cells by either phenol-chloroform or silica column-based methods (DNeasy, Zymo IIIC and Zymo IC). Extracted DNA was digested with Degradase Plus and analysed by LC-MS. The extent of ion suppression was visualised by the dC/dG ratio.

(b) Quantification of the abundance of dC, measured as the response. All results show the average and standard error of three biological replicates.

Finally, we developed a method to calibrate for any persisting residual ion suppression. Specifically a master mix of stable isotope labelled synthetic nucleosides, obtained from and synthesised by the group of Prof. T. Carell, was spiked into each sample and standard. Since the concentration of labelled nucleoside should be the same in each sample, significant changes in response are indicative of ion suppression. By calculating a calibration curve based on the deviation from the expected concentration (light response / heavy response) vs expected concentration, we can correct for matrix effects (Fig. 3.5c). Using this approach, the labelled nucleosides should calibrate for suppression caused by the concentration of both RNA and DNA nucleosides and any matrix effects caused by the digestion mixture (Fig. 3.5b).

To confirm that this protocol is able to accurately quantify DNA modifications from small number of cells, I extracted DNA from 50, 100, 500 and 1000 mESCs grown in FCS and 2i culture conditions using the Zymo Duet IC columns. The DNA was then digested using the NEB enzyme pre-cleaned using the method described above. Figure 3.6 shows the normalised abundance of dC and 5mdC as an average of two technical replicates. Using glycerol-free enzyme and calibrants, the 5mdC/dG appears to be comparable between all 4 samples (50,

100, 500 and 1000 cells). However, this analysis must be repeated with additional biological replicates to determine if the quantified 5mdC differs significantly between samples.

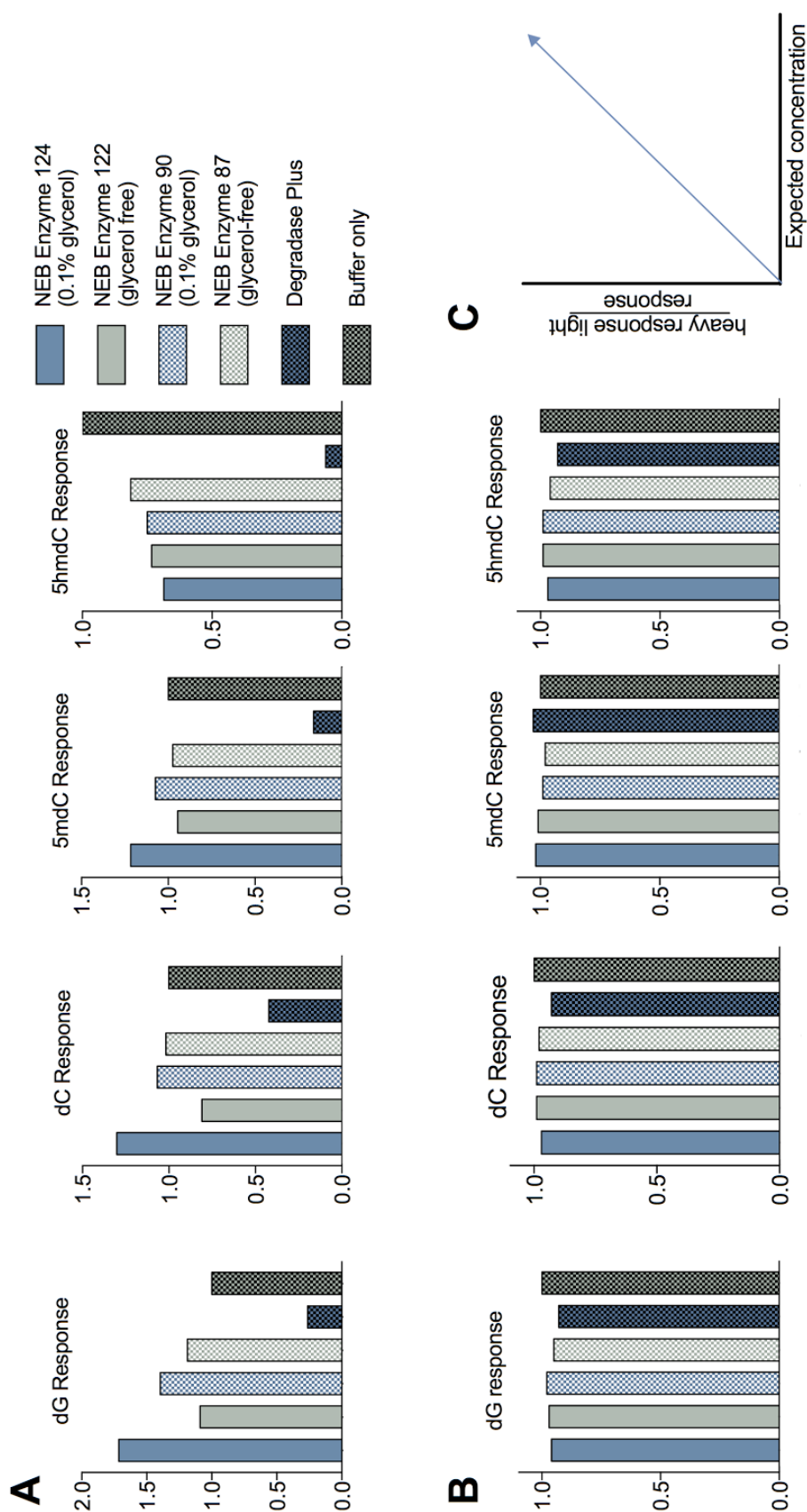


Figure 3.5 Glycerol is a major source of ion suppression

(a) Digestion matrix from 5 different digestion mixtures (and a buffer only control) was added to 250 fmol of synthetic nucleoside standards. The effect of matrix on ion suppression was measured by normalising the response to that of the buffer-only control.

(b) The addition of stable-isotope labelled synthetic standards enables ion suppression caused by the glycerol content of the enzyme matrix to be corrected.

(c) Stable isotope labelled nucleosides enable the effects of ion suppression to be quantified, and a correction factor applied to the measured response.

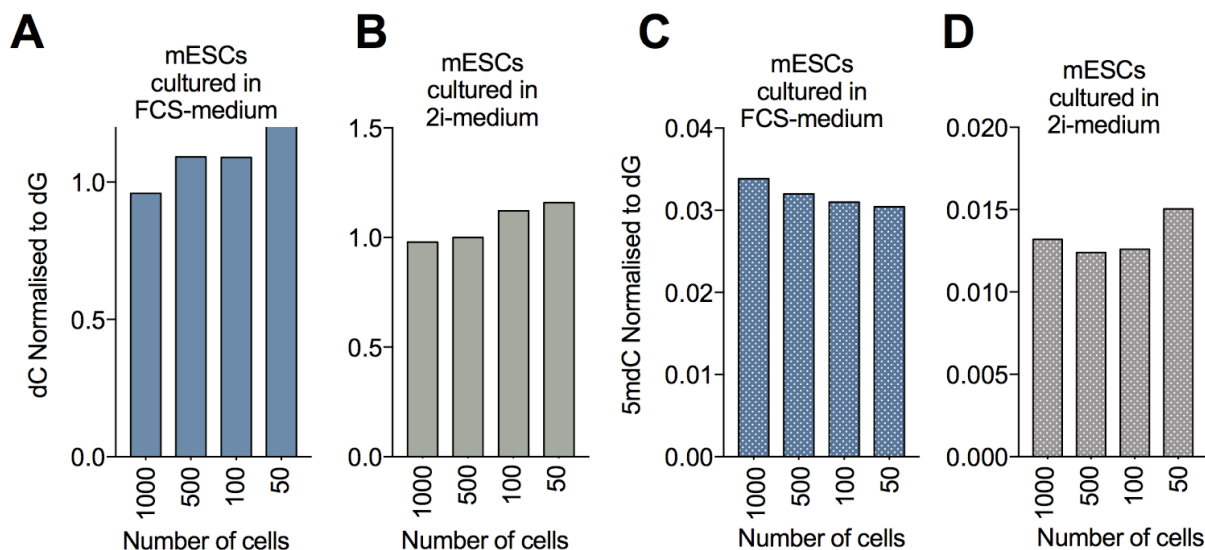


Figure 3.6 Quantification of 5mdC in small cell numbers

mESCs cultured in either FCS- or 2i medium were counted by FACS. DNA was isolated from 1000, 500, 100 and 50 cells from mESCs cultured in each condition. DNA was digested using NEB glycerol free enzyme (No. 122). DNA modifications were quantified by LC-MS using stable isotope labelled nucleosides to correct for any residual ion suppression.

The results shown are the average values from two technical replicates.

(a) Quantification of dC in mESCs cultured in FCS-conditions. Results are normalised to the abundance of dG.

(b) Quantification of dC in mESCs cultured in 2i-conditions. Results are normalised to the abundance of dG.

(c) Quantification of 5mdC in mESCs cultured in FCS-conditions. Results are normalised to the abundance of dG.

(d) Quantification of 5mdC in mESCs cultured in 2i-conditions. Results are normalised to the abundance of dG.

3.3.3 Compatibility of common cytological fixatives with nucleoside LC-MS

Cross-linking agents (including formaldehyde and glutaraldehyde) are the most universally used cell fixatives. However, successful, reproducible extraction of high quality 'fixed' DNA suitable for downstream applications is technically challenging and a potential source of experimental artifacts. For instance, formaldehyde disrupts the intra-chain hydrogen bonding and base stacking; creating sites which are vulnerable to further chemical interactions, including the hydrolysis of N-glycosidic bonds. Covalent modification of nucleic acid bases alter the molecular mass, causing these modified bases to be lost via a collisional or unstable trajectory in a MRM scan. Additionally, DNA yields may be significantly lower than expected due to reduced DNA solubility and the inaccessibility of DNA encased in cross-linked protein complexes.

To assess the compatibility of common cytological fixatives with downstream LC-MS analysis, mESCs were fixed with either paraformaldehyde, zinc-buffered fixative (ZBF) or ethanol. Equal number of cells (100,000) were collected by fluorescent activated cell sorting (FACS). DNA was extracted using the Zymo DNA/RNA duet kit and the DNA yield was quantified. DNA was then digested using our optimised digestion protocol and the nucleosides quantified by LC-MS. dC/dG and 5mdC/dG values from the fixed samples are compared to DNA samples isolated from unfixed cells as a measure of how reliably DNA extracted from the fixed samples can be quantified (Fig. 5.7).

DNA yields from cells fixed using ZBF are comparable to unfixed cells whilst PFA fixation substantially lowers DNA yields (data not shown). Ethanol fixation increases DNA yields compared to unfixed cells, most likely due to DNA precipitation. dC/dG and 5mdC/dG values are comparable between unfixed, ethanol fixed and ZBF fixed cells. However, nucleosides extracted from PFA-fixed cells have a significantly higher dC/dG ratio and a lower 5mdC/dG ratio than expected. Therefore, methods involving cytological fixation prior to downstream LC-MS analysis should avoid the use of cross-linking agents such as PFA.

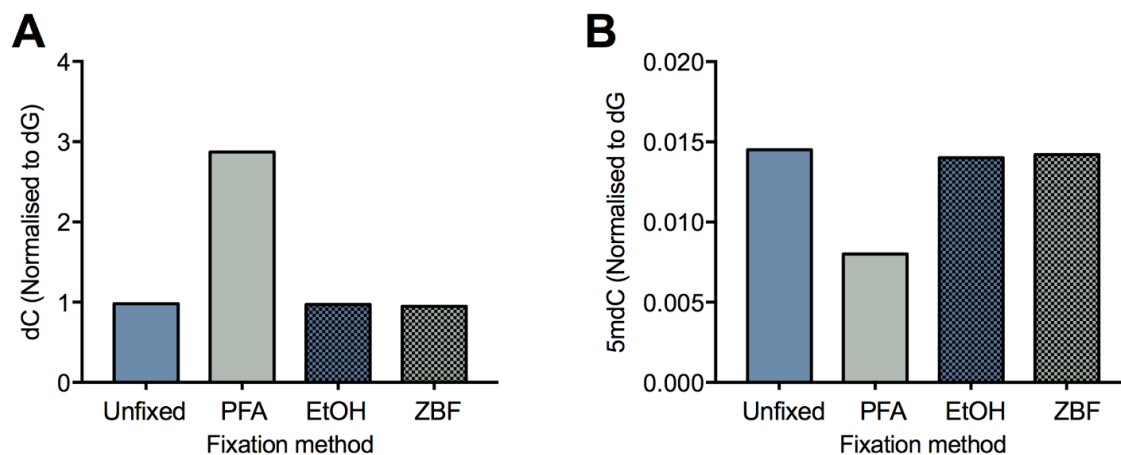


Figure 3.7 Cytological fixation techniques and their effect on LC-MS ion suppression

(a) mESCs cultured in 2i were fixed with 4% PFA for 15 minutes at room temperature, fixed for 12 hours in EtOH or 12 hours in Zinc-buffered fixative. DNA was extracted using the Zymo DNA/RNA Duet DNA miniprep kit with IIIC columns and was digested using our purified NEB digestion mix. dC/dG ratios were calculated as a measure of ion suppression.

(b) 5mdC abundance was also quantified.

3.4 Discussion

The validation of any LC-MS method is vital to assess the accuracy and reliability of a given method. In this instance, accuracy is a measure of how close a measurement is to the actual concentration whilst reliability is a measurement of robustness or resistance to variation. The utility of our initial LC-MS methods were limited by their accuracy; frequently dC/dG ratio deviated from the expected ratio of 1. Additionally, the sample preparation procedure introduced a substantial amount of variability. Ion suppression may manifest as both reduced analytical accuracy and technical reproducibility. Ideally, sample preparation would involve the purification of the sample to remove compounds which are likely to result in matrix effects including ion suppression.

There are numerous potential sources of ion suppression and preventing their deleterious effects is a vital step in the development of any LC-MS method. The results presented in this chapter describe several of the methods we now routinely use to reduce ion suppression including sample dilution, the removal of possible interferents from the digestion matrix, and calibration using isotopically labelled nucleoside standards. The combination of methods used depends on the nature of the samples and particularly on the abundance of analytes of interest. Ion suppression is particularly troublesome for the quantification of DNA modifications in small numbers of cells where the concentration of nucleosides is low relative to the amount of interferents originating from exogenous sources including the digestion matrix. Ion suppression is typically evaluated by monitoring the detector response following the continuous, post-column infusion of potential interferents. Internal standards, including the ratio of dG and dC, are also a useful measure of the extent of ion suppression. The analysis of DNA modifications in small number of mammalian cells, including oocytes, zygotes, early embryos and primordial germ cells, is effectively overcome using purified, glycerol-free digestion enzymes and stable isotope labelled calibrants.

The development of sensitive and quantitative LC-MS technologies has greatly improved our understanding of global epigenetic reprogramming events. Using the described methods, it is possible to quantify the abundance of 5mdC and 5hmdC in just 50 cells. This sensitivity is unprecedented. Our group has used these methods to interrogate the dynamics of global epigenetic reprogramming in the zygote and primordial germ cells (Amouroux et al., 2016)(Hill et al., publication pending). The findings from these two studies oppose the popular idea that global active DNA demethylation is triggered by TET protein driven 5mdC hydroxylation.

In our LC-MS system, the sensitive quantification of nucleosides relies on the efficient ionisation and nebulisation of analytes eluted from the HPLC column. Many compounds, including aldehydes and carboxylic acids, are notoriously difficult to nebulise and therefore the sensitivity of their detection is typically low. The ionisation efficiency of 5fC and 5CaC is low. Since both modified bases are transient and present at very low levels *in vivo*, low sensitivity is particularly problematic for these species. In these scenarios, compounds may be chemically derivatised to groups which are more readily ionised (Gao et al., 2005)(Halket et al., 2006)(Johnson et al., 2007). For instance, Girard's reagents (GirD, GirP and GirT) contain a readily ionisable moiety and react with carboxylic acids to form a stable hydrazone. Future experiments requiring the quantification of these bases may benefit from the development of LC-MS methods involving an additional derivatization step. Importantly, LC-MS is a highly versatile tool that can be adapted to meet the needs of the rapidly evolving field of epigenetics.

Chapter 4

In vivo and *ex vivo* DNA methylation turnover
kinetics

4.1 Introduction

4.1.1 Dynamic DNA methylation

Most cells in the mammalian body are non-dividing and exit the cell cycle permanently. Therefore, the majority of cell types (excluding primordial germ cells and cells of the preimplantation embryo) are believed to have relatively stable DNA methylomes. However, there is a growing list of examples of DNA methylation changes in terminally differentiated, somatic cells; suggesting that DNA methylation may be far more dynamic than previously appreciated (Guo et al., 2011)(Yamagata et al., 2011). Standard means of quantifying DNA methylation are unable to distinguish between static and dynamic methylation. Instead, most methods measure 'steady state' 5mdC levels generated by opposing DNA methylating and demethylation activities.

Dynamic DNA methylation can be analysed using stable isotope labelling and mass spectrometry. Specifically, the use of isotopically labelled methionine, the metabolic precursor for S-adenosylmethionine (SAM), and isotopically labelled uridine can label the 5' methyl group of 5mdC and the pyrimidine ring respectively (Fig.4.1)(Herring et al., 2009). The mass of these precursors is additive. Thus an ensemble of isotopomers is generated; ultimately enabling *de novo* or *maintenance* methylation to be distinguished.

A

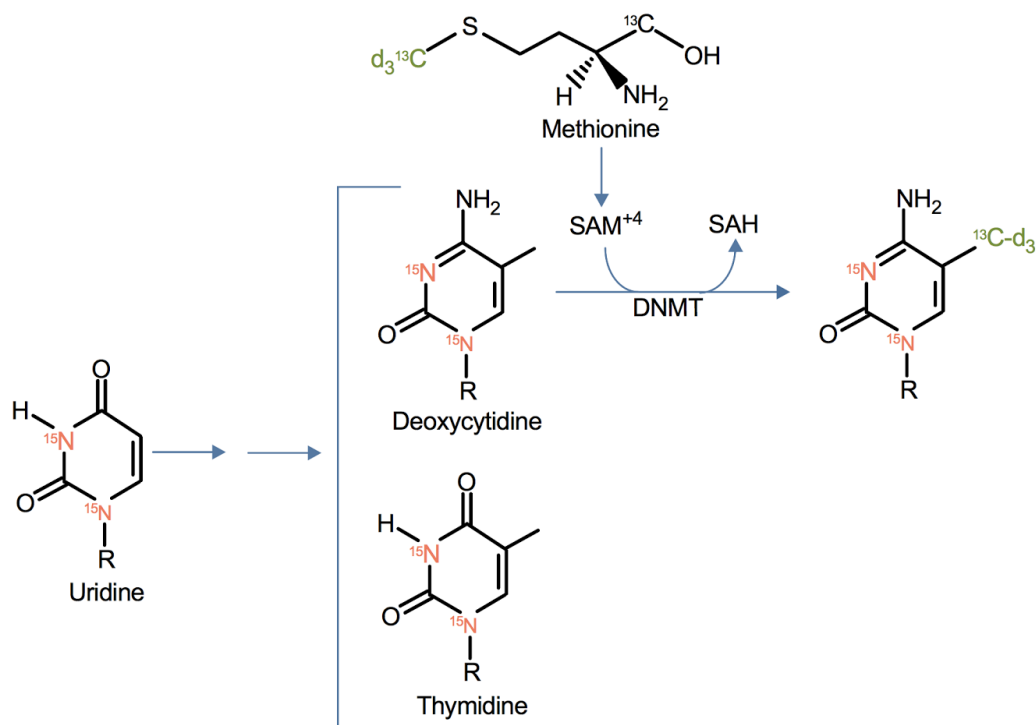


Figure 4.1 Stable isotope labelling of pyrimidines and 5-methylcytosine

(a) Schematic showing how the pyrimidine ring and the methyl group on C5 can be labelled via stable isotopically labelled uridine ($^{15}\text{N}_2$ -Uridine) and stable isotopically labelled L-methionine ($^{13}\text{C-d}_3$ L-methionine) respectively.

4.1.2 *De novo* pathway of pyrimidine biosynthesis

The development of an experimental system that relies on labelled uridine as a source of pyrimidines requires an understanding of pyrimidine biosynthesis. In resting cells, nucleotides (both deoxy- and ribonucleotides) are thought to account for 15-20% of a cell's biomass (Lane and Fan et al., 2015). Since RNA is typically transcribed as large precursor molecules which are subsequently spliced, 3-fold more RNA synthesis is required than the amount of mature RNA present in a given cell (Lane and Fan et al., 2015). Therefore, the vast majority (12-21%) of these nucleotides are assembled into RNA (Alberts et al., 1994). Prior to cell division, the cell's entire biomass must double, meaning that nucleotides are constantly synthesised in all cells (*de novo* synthesis). In addition, nucleotides released via hydrolysis of polynucleotides can be recycled via the salvage pathway (Fig. 4.2). Concentrations of nucleotides are typically maintained within narrow limits. Cytosolic concentrations of

nucleosides have been measured in the order ATP > UTP > GTP > CTP (Traut, 1994). Therefore, nucleotide biosynthesis must be highly regulated to not only take into account the rate of cell proliferation but also the amount of RNA catabolism.

The relative physiological importance of the *de novo* and salvage pathways depends on the cell type and its developmental stage. *De novo* synthesis is indispensable for highly metabolically active proliferating cells. However, the *de novo* pathway activity is typically low in differentiated cells (Fairbanks et al., 1995). *De novo* pyrimidine synthesis is regulated at the transcriptional level by a set of master regulators. These biosynthetic enzymes are subject to allosteric inhibition, activation and post-translational modifications. Of these mechanisms, feedback inhibition of the first committed step by a product of the final step is widespread (Fig. 4.2)(Reviewed in Jones et al., 1980)(Evans et al., 2004).

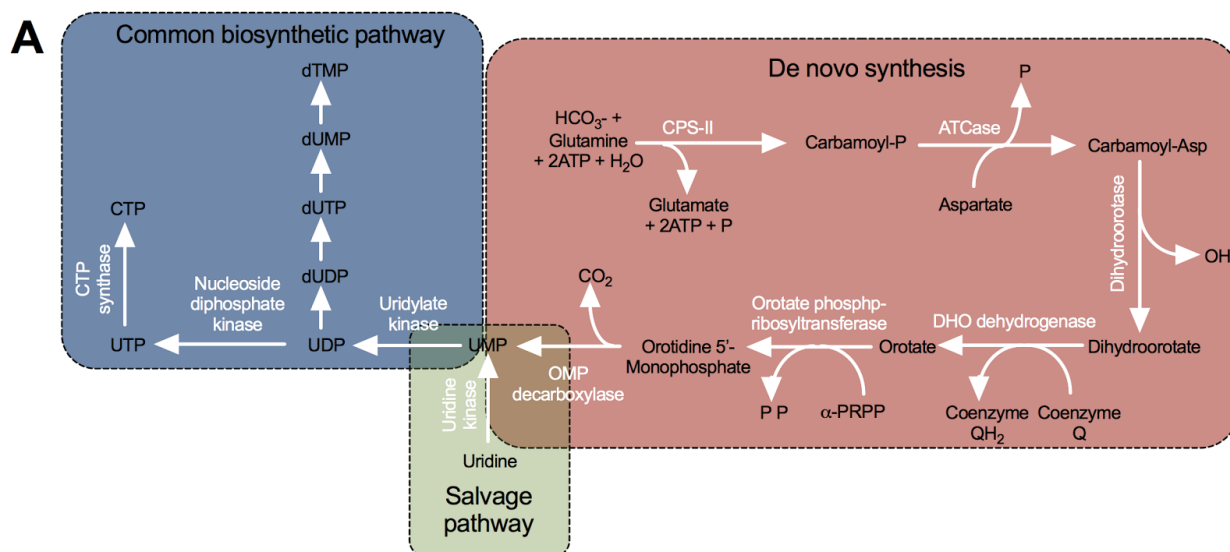


Figure 4.2 Pathways of pyrimidine biosynthesis

In the first step of *de novo* pyrimidine synthesis, carbamoyl phosphate is produced by the phosphorylation of glutamine (Fig. 4.2). This reaction, catalysed by the carbomyl phosphate synthetase II activity of the trifunctional enzyme CAD (Coleman et al., 1977), is rate limiting in pyrimidine biosynthesis; being subject to

feedback inhibition by cytosolic uridine diphosphate (UDP) and uridine triphosphate (UTP) and activated by 5-phosphoribosyl-1-pyrophosphate (PRPP), an activated form of ribose (Jones et al., 1980).

Aspartate transcarbamoylase, the second enzymatic activity of CAD, catalyses the condensation between carbamoyl phosphate and aspartic acid; generating carbamoyl aspartate. Circularisation of carbamoyl aspartate, catalysed by the final enzymatic activity of CAD (dihydroorotase), produces dihydroorotate. Dihydroorotase dehydrogenase (Dhohd) subsequently oxidises dihydroorotate using NAD^+ . Thus, pyrimidine biosynthesis is intricately linked to cellular metabolism; reduced *de novo* pyrimidine synthesis is known to prolong the cell cycle under oxygen limited conditions (Sigoillot et al., 2002). Dihydroorotate oxidation generates the base orotate. Subsequent coupling of orotate to the activated ribose, PRPP, produces the pyrimidine orotidylate (orotidine 5-monophosphate, OMP). Orotidylate is then decarboxylated; producing uridylate (UMP) (Suttle et al, 1988). UMP, the precursor for all pyrimidine biosynthesis, is converted into an active form (di- and triphosphates) by specific nucleoside monophosphate kinases (UMP kinase) and nucleotide diphosphate kinases before conversion into either CTP or TTP.

4.1.3 Salvage pathway of pyrimidine biosynthesis

The salvage pathway essentially consists of importing exogenous bases into the cell and recycling nucleic acid degradation products. Unlike purine salvage, pyrimidine are recycled from their nucleosides rather than their bases. This is evidenced by the fact that symptoms associated with deficient *de novo* pyrimidine synthesis can be overcome by the administration of exogenous uridine, but not uracil. The salvage pathway of pyrimidine biosynthesis is highly energy efficient; a single mole of ATP is required to produce a single mole of pyrimidine nucleotide compared to 5 mole of ATP to produce a mole of pyrimidine via the *de novo* pathway. Although proliferating cells appear to preferentially rely on the *de novo* pathway of pyrimidine synthesis, rates of *de novo* synthesis are reduced when exogenous uridine is supplied (Hoogenraad and Lee, 1974). Additionally, disruption of *de novo* synthesis (for example, using Pyrazofurin) imposes an absolute nutritional requirement on the utilisation of preformed bases without being lethal (Cadman et al., 1980).

Pyrimidine salvage via uridine is catalysed by uridine monophosphate synthase, a uridine kinase (UDK) (Fig. 4.2). Both UTP and CTP are inhibitors of this reaction. These kinases are sensitive to the levels of accumulating biosynthetic intermediates such that UTP accumulation inhibits UMP kinase, instead directing UMP towards byproduct formation and excretion as uracil.

4.2 Experimental aims

We questioned the paradigm that DNA methylation is a stable modification. Since, DNA methyltransferases and proteins involved in active DNA (de)methylation are expressed in many postmitotic cells, and transcription and DNA repair are known to disrupt the underlying methylation state, we reasoned that DNA methylation could potentially be highly dynamic.

Therefore, my first aim was to develop a method to quantify dynamic 'steady state' DNA modifications using stable isotope labelling and LC-MS. I then wanted to employ this method to assess the dynamics of 5mdC and its oxidative derivatives in the absence of DNA replication (postmitotic cells). This required the development of a suitable model system. Specifically, any cells used for this experiment had to be non-replicative and terminally differentiated. Additionally, the collected cell population had to be homogeneous. Finally, I wanted to use knockout (or knockdown) models to interrogate the role of putative DNA demethylases in DNA methylation turnover using a mESC culture system.

4.3 Results

4.3.1 Experimental system

In both *in vivo* and *ex vivo* experimental systems, newly methylated dC can be isotopically labelled. For example, since the universal methyl donor, S-adenosylmethionine (SAM), is a conjugate of methionine and the adenosine moiety of ATP, the addition of ^{13}C - d_3 -enriched methionine to animal feed or culture medium will generate ^{13}C - d_3 labelled 5mdC (Fig. 4.1). In proliferating cell populations, ^{13}C - d_3 5mdC would be abundant due to replication-dependent maintenance methylation.

To assess the dynamic ‘steady-state’ of DNA modifications, it is important to identify all ^{13}C - d_3 5mdC that is the result of replication-independent (*de novo*) DNA methylation. The degree of DNA replication can simultaneously be quantified using stable isotope labelled pyrimidines. Specifically, the pyrimidine precursor pool can be labelled by supplementing culture medium with $^{15}\text{N}_2$ -enriched uridine (Fig. 4.1). In this system, the efficient biosynthesis of $^{15}\text{N}_2$ -dCTP and $^{15}\text{N}_2$ -dTTP relies on the use of $^{15}\text{N}_2$ -uridine as the primary source of UMP (the salvage pathway) (Fig. 4.2). Ultimately, an ensemble of 5mdC isomers differing in the degree of isotopic enrichment (isotopomers) is generated (Table 4.1 & Fig. 4.3). These isotopomers can be individually quantified by LC-MS. By comparing the rate constants for the incorporation of each isotopomer, it may be possible to determine whether ‘steady state’ DNA methylation conceals the more dynamic nature of the DNA methylome.

4.3.2. Detection of isotopomers

Several adaptations had to be made to the LC-MS method described in Chapter 3 to enable *de novo* and maintenance methylation to be quantified based on the abundance of isotopomers. First and foremost, an acquisition method was required to detect and quantify each isotopomer of interest. This process involves

optimising the collision energy required to fragment the precursor ion (i.e. break the N-glycosidic bond), optimising the dwell time (the time devoted to detecting a specific m/z transition during each cycle) and ensuring that the expected ratios of molecular adducts (H^+ , K^+ and Na^+) are obtained. Typically one should use synthetic, high purity analytes when developing an LC-MS acquisition method. In this instance, however, the isotopomers of interest are not commercially available. Instead, I developed my own mESC culture based system to generate biological samples enriched for each isotopomer.

E14 (129/Ola) mESCs were grown in one of four culture conditions for 72 hours (Fig. 4.3a and Fig. 4.3b):

1. unlabelled 'light' FCS-medium;
2. FCS-medium supplemented with $^{15}N_2$ -uridine;
3. FCS-medium with ^{13}C -d₃ L-methionine as the sole exogenous source of SAM; or
4. FCS-medium supplemented with $^{15}N_2$ -uridine *and* using ^{13}C -d₃ L-methionine as the only exogenous source of SAM.

Following DNA isolation and digestion, diluted nucleosides were analysed by LC-MS. Each isotopomer is quantified using a standard curve of 'light', un-enriched nucleosides. The quantification thus relies on the assumption that both the isotopically labelled and 'light' nucleosides behave identically in terms of their retention time on the HPLC column, electrospray ionisation properties and their fragmentation pattern during CID. We have determined that this is the case.

As expected, isotopically-enriched ($^{15}N_2$) dT can only be detected in cells grown in FCS-medium supplemented with $^{15}N_2$ -uridine (culture conditions 2 and 4)(Fig. 4.3b). Similarly, ^{13}C -d₃ labelled isotopomers are only produced in FCS-medium containing ^{13}C -d₃ L-methionine (culture conditions 3 and 4). Interestingly, the enrichment of $^{15}N_2$ dT is higher in mESCs cultured in condition 4 (^{13}C -d₃ L-methionine & $^{15}N_2$ -uridine) compared to condition 2 ($^{15}N_2$ -uridine alone), possibly indicating that the extracellular concentration of L-methionine is a key determinant of cell proliferation rate and DNA synthesis. After 72 hours in culture condition 2 and 4, 30% and 50% of genomic dT is isotopically labelled respectively. The limited incorporation of

$^{15}\text{N}_2$ dTTP and dCTP into the genome may be explained by inefficient $^{15}\text{N}_2$ -uridine salvage and the continued reliance on the *de novo* pathway of pyrimidine biosynthesis..

In contrast, replacing all 'light' L-methionine with ^{13}C -d₃ L-methionine (culture condition 3) generates an *almost* homogeneous population of 5mdC isotopomers (C-Me*, where * denotes whether the pyrimidine ring or methyl group is isotopically labelled). Therefore, perhaps unsurprisingly, the transfer of the isotopically labelled methyl group from ^{13}C -d₃ SAM appears to be far more efficient than the labelling of the pyrimidine precursor pool via $^{15}\text{N}_2$ -uridine. Importantly however, between the four culture conditions, we can detect all four 5mdC isotopomers of interest: C-Me, C*-Me, C-Me* and C*-Me*. Additionally, the sensitivity of detection for dC* and dT* appears to be high.

To increase the versatility of this method (i.e. by enabling the quantification of isotopomers in genomic DNA from a wide range of biological samples), it is important to ensure that we can calibrate for the effects of ion suppression. As described in Chapter 3, in many instances, the deleterious effects of ion suppression can be resolved using stable isotope-labelled nucleosides (calibrants). It is therefore important to ensure that the calibrants and isotopomers generated via *in vivo* or *ex vivo* metabolic labelling do not have the same m/z transition. Table 4.1 shows the m/z transition of each labelled nucleotide we expect to detect. Importantly, neither the calibrants nor the $^{15}\text{N}_2$ -uridine and ^{13}C -d₃ L-methionine are 100% pure. Therefore, a small proportion of the compounds are likely to be either partially labelled or unlabelled. Appropriate controls must therefore be used to ensure that genomic isotopomers are indeed the result of metabolic labelling. For these negative controls, I used unlabelled sample spiked with the calibrants described in Chapter 3 and labelled samples without the calibrant.

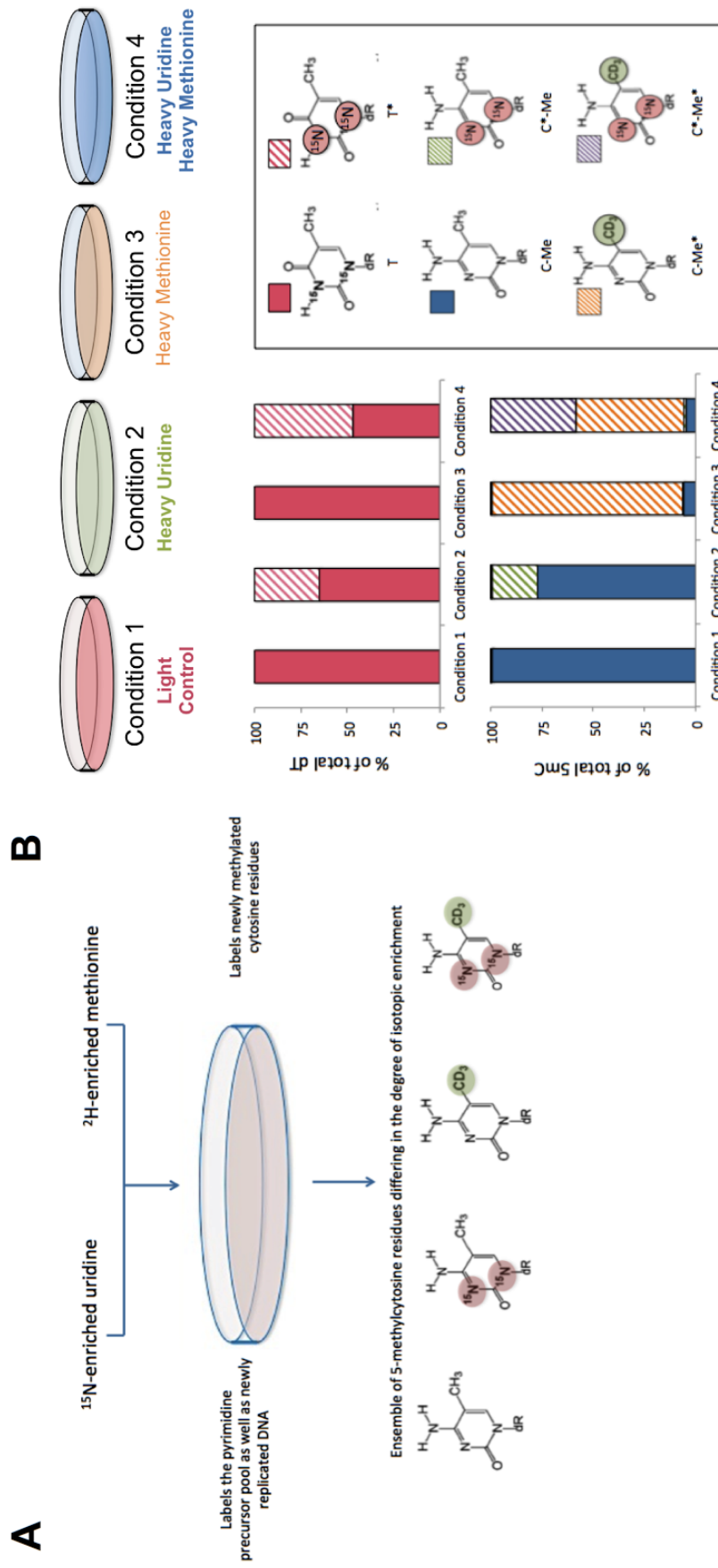


Figure 4.3 Validation of LC-MS method to detect individual isotopomers

Table 4.1: Summary of additional compound-dependent parameters

Species	Precursor m/z	Product m/z	CE	Dwell time	Cell accelerator voltage	Δ EMV
dT H	243.1	127.1	10	15	5	300+
dT Na	265.1	149.1	10	100	5	300+
dT K	249.1	133.1	10	15	5	300+
dT* H	245.1	129.1	10	15	5	300+
dT* Na	267.1	151.1	10	100	5	300+
dT* K	283.1	167.1	10	15	5	300+
C-Me* H	246.1	130.1	10	30	5	400+
C-Me* Na	268.1	152.1	10	100	5	400+
C-Me* K	284.1	168.1	10	30	5	400+
C*-Me H	244.1	128.1	10	30	5	400+
C*-Me Na	266.1	150.1	10	100	5	400+
C*-Me K	282.1	166.1	10	30	5	400+
C*-Me* H	248.1	132.1	10	30	5	400+
C*-Me* Na	270.1	152.1	10	100	5	400+
C*-Me* K	286.1	170.1	10	30	5	400+
dC* H	230.1	114.1	8	30	5	500+
dC* Na	252.1	136.1	8	100	5	500+
dC* K	268.1	152.1	8	30	5	500+

4.3.3. Stable isotope labelling and quantification of 5mdC turnover in postmitotic cells

Using ^{13}C -d₃ L-methionine, we wanted assess whether 5mdC is cyclically removed and replaced in terminally differentiated, postmitotic cells. We first turned our attention to postmitotic mouse neurons. Although the adult mammalian brain is *almost* universally regarded as postmitotic, specific examples of localised proliferation and differentiation of adult neural stem cells do exist (Sailor et al., 2016)(Fig 4.4a). The vast majority of this

neurogenesis occurs in the ventricular zone (a layer within the lumen of the neural tube) and later in the subventricular zone (SVZ) of the lateral ventricles and the subgranular zone (SGZ) of the dentate gyrus within the hippocampus. Neurons produced in the SVZ migrate to the olfactory bulb via the rostral migratory stream whereas neurons newly produced in the SGZ are destined to become dentate granule cells (Fig. 4.4a). Whether neurogenesis occurs in other areas of the adult mammalian brain is highly controversial.

To address whether DNA methylation is static or dynamic in postmitotic neurons, we fed adult male mice a custom diet comprising ^{13}C -d₃ stable isotope labelled methionine. After 5, 7 and 10 weeks, the cerebral cortex was isolated and homogenised. Importantly, regions believed to be the sites of neurogenesis were carefully avoided. We decided to use a FACS approach to separate neuronal and non-neuronal (i.e. glial and microvasculature) cells. However, FACS requires a suspension of cells of fairly uniform size to reliably gate for doublets and signal intensity. The cells of the cerebral cortex cannot be isolated intact and are highly variable in terms of size. Therefore, I first isolated nuclei from each cerebral cortex to facilitate the separation of postmitotic cortical neurons and other cell types.

Criteria for cell lysis and nuclear isolation were optimised to produce clean and reproducible nuclei preparations (Fig 4.4b). In particular, the osmolarity and volume of extraction buffer had significant impact on the efficiency of cell lysis. Hypotonic cells were lysed using a Dounce homogeniser and contaminating cytoplasm, cell membrane and myelin sheath was removed by repeated washing. The purity of the prep and nuclei integrity was determined by staining with trypan blue. Typically, nuclei should be round or oval with uniform dark blue staining. Any remaining cytoplasm clearly stains a lighter shade of blue. Intact cells remain unstained. Unfixed, isolated nuclei were then labelled using NeuN, a marker of postmitotic neurons. NeuN positive and negative nuclei were collected separately by FANS (fluorescent activated nuclei sorting) (Fig. 4.4c). A low flow rate was chosen to limit accidental lysis. Nuclei were then directly lysed in lysis buffer for downstream DNA extraction and digestion to nucleosides for LC-MS analysis.

The LC-MS analysis showed that NeuN positive postmitotic neurons are more highly methylated than the remaining cells in the cortex; 3.5% of cytosines in NeuN positive nuclei are methylated compared to 3% of cytosines in the remaining NeuN low nuclei. Figure 4.4d shows the average 5mdC/dG of 4 biological replicates. Error bars show the standard error of the mean (SEM). Using an unpaired t-test, this difference was deemed significant (p-value < 0.05). Interestingly, NeuN positive nuclei also have a remarkably high proportion of hmdC; 0.7% cytosines are hydroxymethylated compared to 0.28% cytosines in NeuN low nuclei. These values are in agreement with previous analysis, suggesting that our applied method is sound (Munzel et al., 2010).

My analysis further showed that the proportion of *in vivo* stable isotope labelled 5mdC is higher in NeuN low nuclei compared to NeuN high nuclei. This is expected given that the NeuN low population includes numerous different cell types, many of which are expected to be proliferative. More surprising is the increase in isotope-labelled 5mdC seen in NeuN high nuclei labelled for progressively longer periods; 0.005% 5mdC after 5 weeks to just less than 0.02% 5mdC at 10 weeks. Possible explanations for stable isotope labelling of neurons include:

- the addition of a labelled methyl moiety to a cytosine via labelled S-adenosyl methionine;
- DNA repair, for example short- or long-patch BER and subsequent methylation of the newly incorporated cytosines;
- contamination with false NeuN positive nuclei; or
- low level neurogenesis.

It will be important to repeat this experiment to determine whether these results are reproducible.

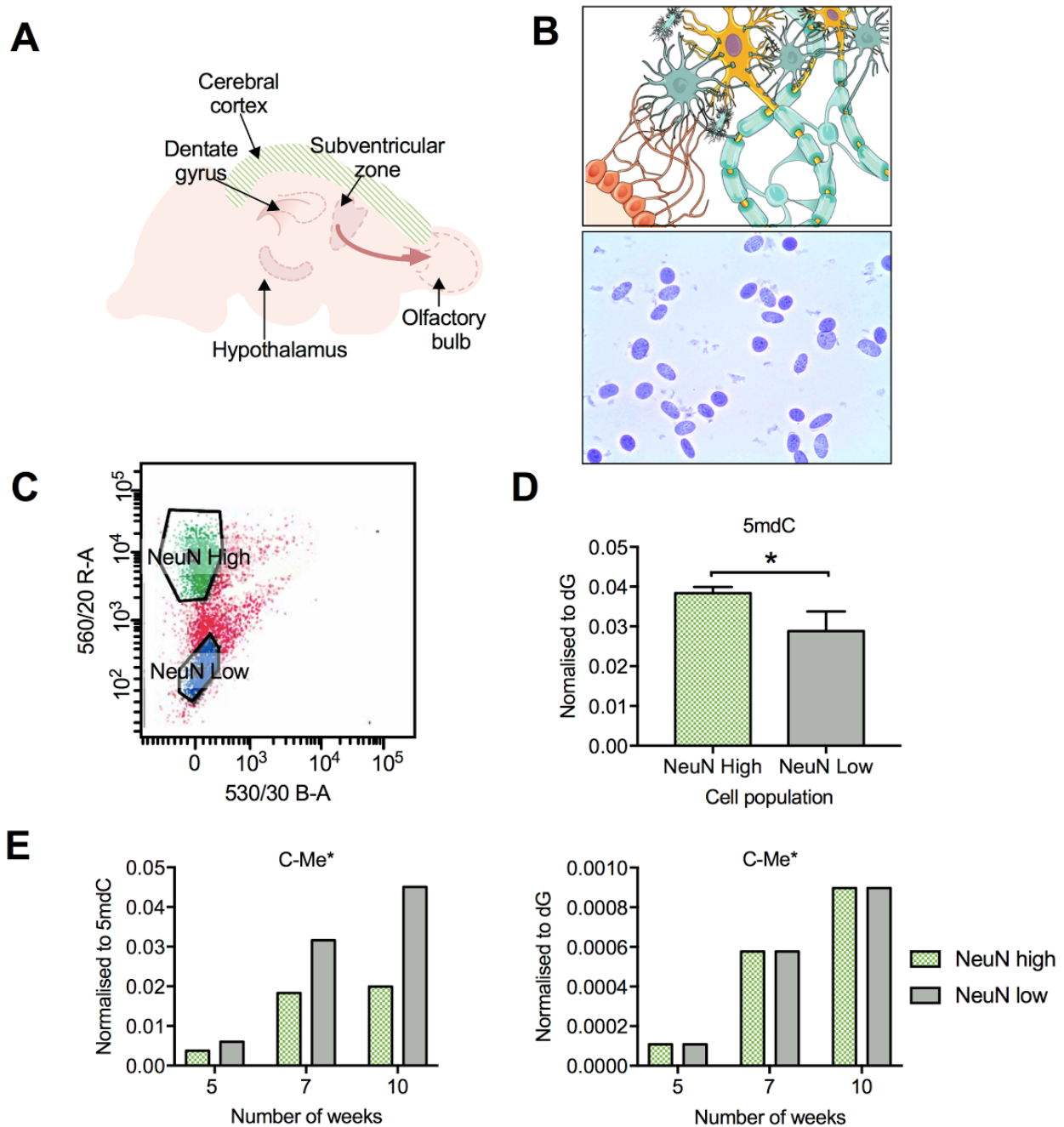


Figure 4.4 *In vivo* ^{13}C -d₃ L-methionine labelling of adult cortical neurons

(a) Schematic illustrating the regions of neurogenesis in the adult mouse brain and the location of tissue harvested for LC-MS analysis (green hashed).
 (b) Trypan blue staining of extracted nuclei following optimisation of the nuclei extraction protocol.
 (c) FACS profile illustrating the gating strategy used to isolate populations of NeuN high and NeuN low nuclei.
 (d) LC-MS analysis of NeuN high and NeuN low cells from 4 biological replicates. The mean total 5mdC content of NeuN high cells is greater than NeuN-low cells. This difference is statistically significant (unpaired t-test, where * denotes a p-value of < 0.05).
 (e) LC-MS analysis of C-Me* incorporation into cortical cells. Mice were fed a diet enriched in ^{13}C -d₃ for 5, 7 and 10 weeks. DNA was extracted from cortical neurons (NeuN high) and non-neuronal (NeuN low) cells from a single mouse at each timepoint. Each sample was analysed in duplicate and normalised to 5mdC and dG respectively.

Additionally, it may be worthwhile repeating this analysis using BrdU labelled mice to determine whether NeuN exclusively labels a non-proliferative neuronal cell population. Dual labelling could be confirmed using immunofluorescence and confocal microscopy to reduce the likelihood of falsely identifying satellite nuclei (doublets) as a single double labelled nucleus. BrdU can also be quantified using our LC-MS methods with some adjustment to the mass of the precursor and product ions. Comparable levels of isotope labelled 5mdC and BrdU would strongly suggest that any measured 'heavy' 5mdC was the result of cell proliferation; either neurogenesis or contaminating non-neuronal cells.

4.3.4. Optimisation of $^{15}\text{N}_2$ -uridine concentration in proliferating mESCs

Cultured pluripotent mESCs are characterised by dynamic states of DNA methylation. These states can be altered depending on the chosen cell culture medium (FCS or 2i condition)(Leitch et al., 2013)(Ficz et al., 2013)((Habibi et al., 2013). Although the precise mechanism behind culture-induced changes of 5mdC levels is unknown, it is likely to reflect differences in the expression of DNA methyltransferases and factors believed to play a role in DNA demethylation. It is also possible that the dynamics of 5mdC turnover in the genome of mESCs maintained in each of these culture conditions differs. Many mESC lines with deletion of genes involved in DNA (de)methylation have already been generated and characterised (Li et al., 1992)(Dawlaty et al., 2014)(Dr. H. Leitch, personal communication). Therefore, using genetically modified mESCs and the method illustrated in Figure 4.1 and Figure 4.3a, it may be possible to assess the relative contribution of each protein to maintaining the dynamic steady state of DNA modifications in mESCs.

The fact that mESCs are proliferative presents some additional challenges. Particularly since most isotopomers will be incorporated into the genome during DNA synthesis. Determining which isotopomers are the result of maintenance or *de novo* methylation therefore crucially depends on the efficient salvage of $^{15}\text{N}_2$ uridine and the incorporation of dC and dT isotopomers. Pyrimidine biosynthesis in mESCs primarily relies on the *de novo* biosynthetic pathway. The addition of an exogenous source of $^{15}\text{N}_2$ uridine, however is expected to trigger a switch towards the more energy efficient pyrimidine salvage pathway (Hoogenraad and Lee, 1974). Therefore,

higher concentrations of $^{15}\text{N}_2$ uridine may favour the use of the salvage pathway. On the other hand, UMP Kinase, is inhibited by high levels of UTP (a precursor in the biosynthesis of dCTP)(Serina et al., 1995). Above a given threshold, additional UMP will instead be directed towards the cells 'overflow metabolism'; thus, increasing the extracellular concentration of $^{15}\text{N}_2$ uridine will not necessarily result in increased labelled dCTP and dTTP incorporation.

To determine the optimal $^{15}\text{N}_2$ uridine concentration, mESCs derived from embryos of C57B6 genetic background were grown in FCS culture media containing 5 μM , 25 μM , 50 μM , 75 μM and 100 μM $^{15}\text{N}_2$ uridine. The medium was replaced every 12 hours so as to maintain a gradient between the extracellular and intracellular uridine concentrations. Cells were collected every 12 hours for 72 hours. DNA extracted from the cells was digested using our purified, custom made NEB enzyme mixture. The nucleosides were then quantified by LC-MS. There is a strong correlation between $^{15}\text{N}_2$ uridine concentration and the incorporation of labelled pyrimidines. Maximal incorporation of isotopically labelled pyrimidines occurred when using 100 μM $^{15}\text{N}_2$ uridine. Therefore, we decided to use 100 μM $^{15}\text{N}_2$ uridine for the labelling of mESCs.

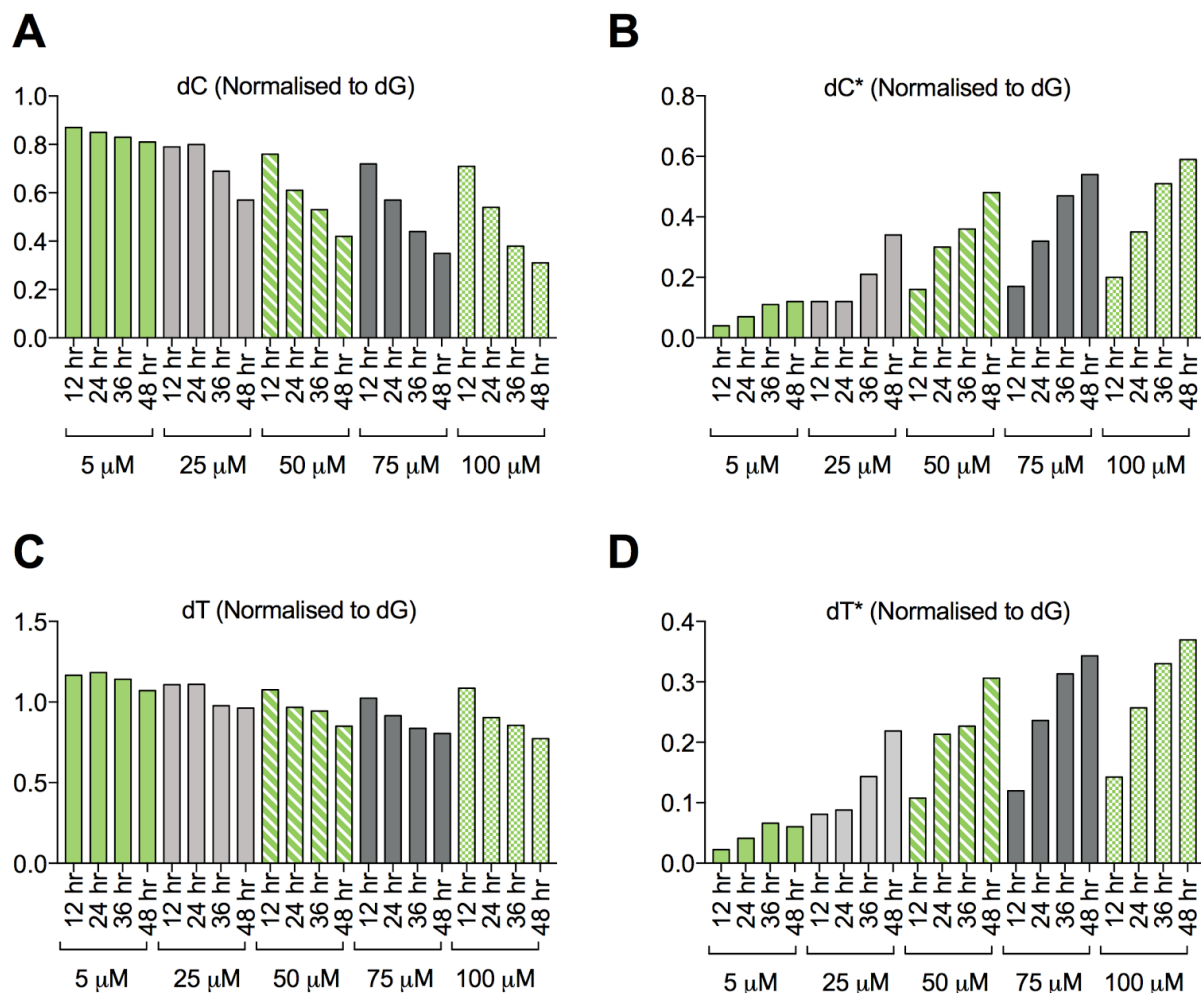


Figure 4.5 $^{15}\text{N}_2$ -uridine concentration and the incorporation of $^{15}\text{N}_2$ -dCTP and $^5\text{N}_2$ -dTTP

C57B6 mESCs were cultured in the presence of $^{15}\text{N}_2$ -uridine at a concentration of either 5 μM , 25 μM , 50 μM , 75 μM , or 100 μM . Cells were collected every 12 hours for 48 hours and the proportion of dC (a), dC* (b), dT (c), and dT* (d) was quantified by LC-MS. The abundance of each isotopomer is normalised to dG. Each sample was analysed in triplicate.

4.3.5. Quantification of isotopomer incorporation in *Tet* double knockout mESCs

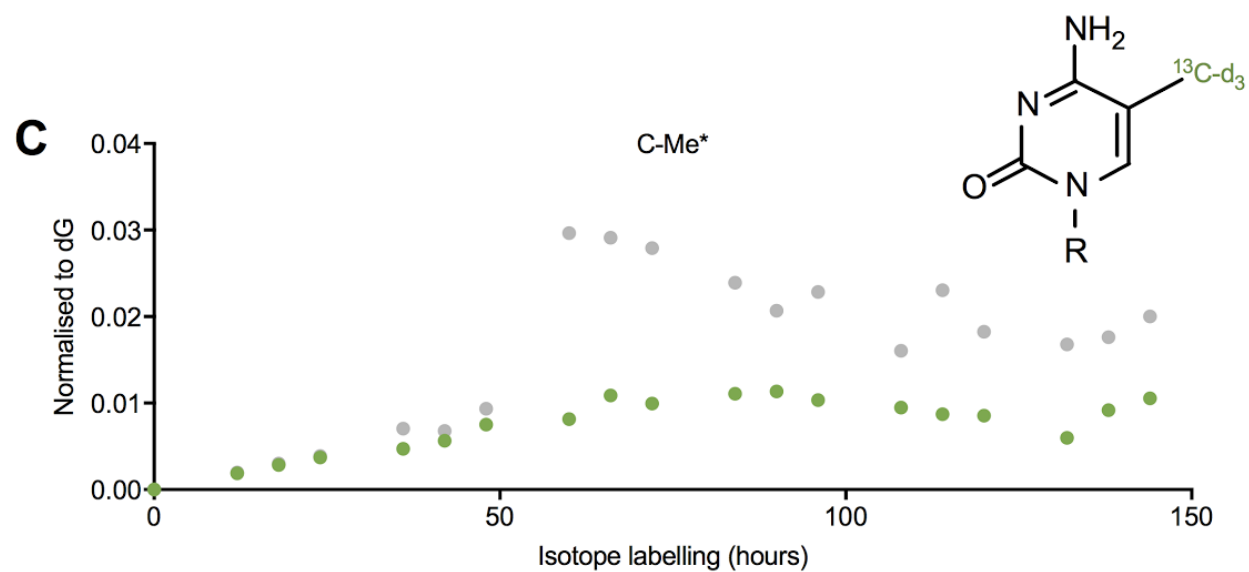
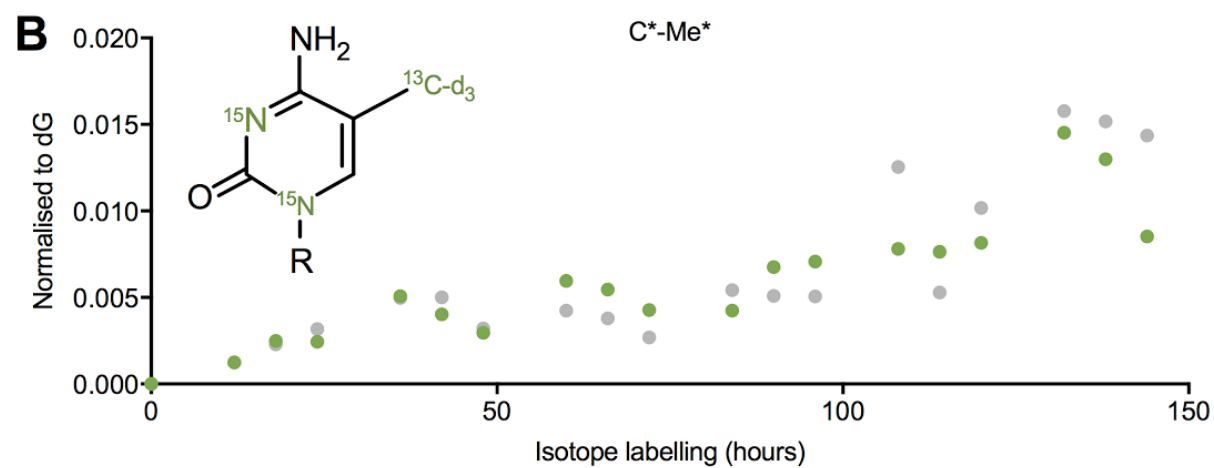
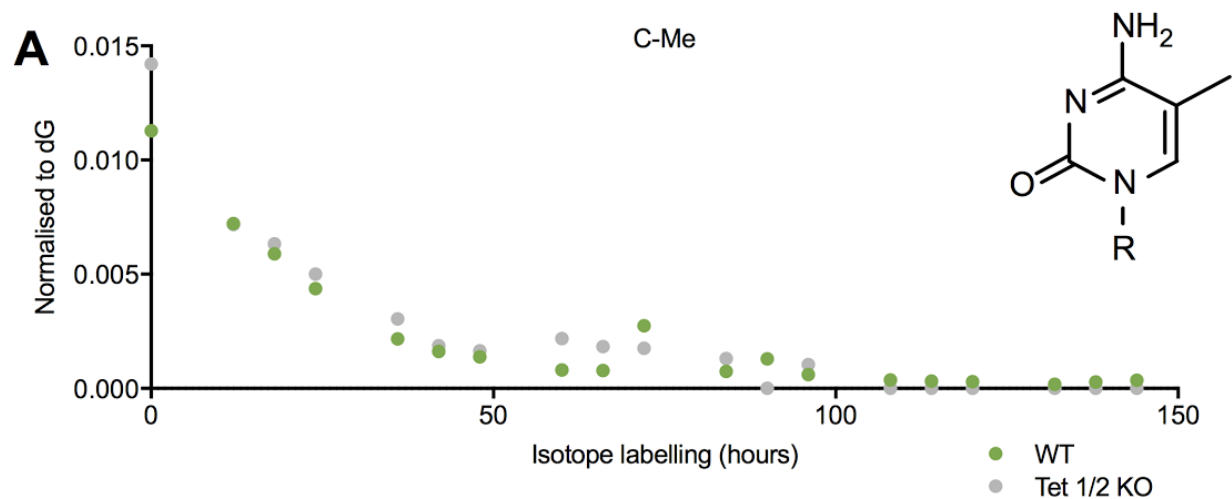
mESCs express high levels of TET1 and TET2 (Ito et al., 2010). To assess the combined role of TET1 and TET2 in dynamic 5mdC removal in a steady state system, we decided to compare the rate of isotopomer incorporation in *Tet* double knockout (DKO, C57B6 genetic background) mESCs and C57B6 wild type mESCs. In proliferating cells, the rate of isotopomer incorporation is expected to scale with the rate of cell proliferation. Therefore, in

this experimental system, labelling of dT and dC must be compared between the cell lines to ensure that the rate of proliferation is not significantly different.

Initial attempts to quantify isotopomers in *Tet* DKO and *Tet* WT mESCs were complicated by the differentiation of mESCs in the isotope labelled 'heavy' medium. Typically, FCS used for the culture of C57B6 mESCs in FCS, feeder-free conditions must be batch tested to ensure that the cells remain pluripotent. Notably, mESCs with other genetic backgrounds appear to be slightly less sensitive to the FCS batch. For stable isotope labelling, dialysed FCS was chosen to reduce the concentration of unlabelled L-methionine and uridine in the culture medium. However, it is likely that this FCS does not support the culture of pluripotent C57B6 mESCs. Furthermore, since neurobasal without L-methionine is not commercially available, conventional 2i conditions are not compatible with our labelling approach. Instead, these C57B6 mESCs can be maintained in a hybrid condition of FCS + ERK and GSK3 β inhibitors. Of course, this is not an ideal system. It is possible that the culture condition have a drastic effect on dynamic 'steady state' DNA methylation. Therefore, it is best to use standard, characterised culture conditions. For future experiments, it may be worthwhile obtaining genetically modified mESCs with a different genetic background (e.g. 129).

To quantify the effect of TET-triggered 5mdC turnover in mESCs, *Tet* DKO and wild type mESCs were cultured in the presence of 100 μ M $^{15}\text{N}_2$ uridine and 6.53 mM ^{13}C -d $_3$ L-methionine 144 hours (Fig 4.8). The proportion of isotopomer incorporation was quantified by LC-MS. When mESCs are cultured in the presence of ^{13}C -d $_3$ L-methionine, 'light' 5mdC is rapidly diluted. The rate of 'light' 5mdC loss is not substantially different between *Tet* DKO and wild type mESCs. However, this may simply reflect the fact that the vast majority of 5mdC turnover in mESCs is likely to occur during DNA synthesis. Quantifying the abundance of C-Me* is a better measure of dynamic 'steady state' DNA methylation. Initially, C-Me* increases at a similar rate in *Tet* DKO and wild type mESCs. From 60 hour onwards, the abundance of C-Me* (and total 5mdC) jumps in *Tet* DKO cells. This is most likely an artifact. Therefore, this experiment must be repeated before any firm conclusions can be drawn.

As described previously, the incorporation of $^{15}\text{N}_2$ -dTTP and $^{15}\text{N}_2$ -dCTP isotopomers is inefficient. This may be due to presence of unlabelled pyrimidines within the culture medium competing with their isotopomers. The rate of pyrimidine isotopomer incorporation is a function of the concentration of $^{15}\text{N}_2$ uridine in the culture medium compared to the concentration within the cells, the rate of synthesis via the salvage pathway and the rate of cell proliferation. To simplify the analysis, it will be informative to culture wild type and *Tet* DKO cells first in medium containing $^{15}\text{N}_2$ uridine and ^{13}C -d₃ L-methionine. After transferring the cells to isotopomer-free medium, the rate of isotopomer loss can be quantified.



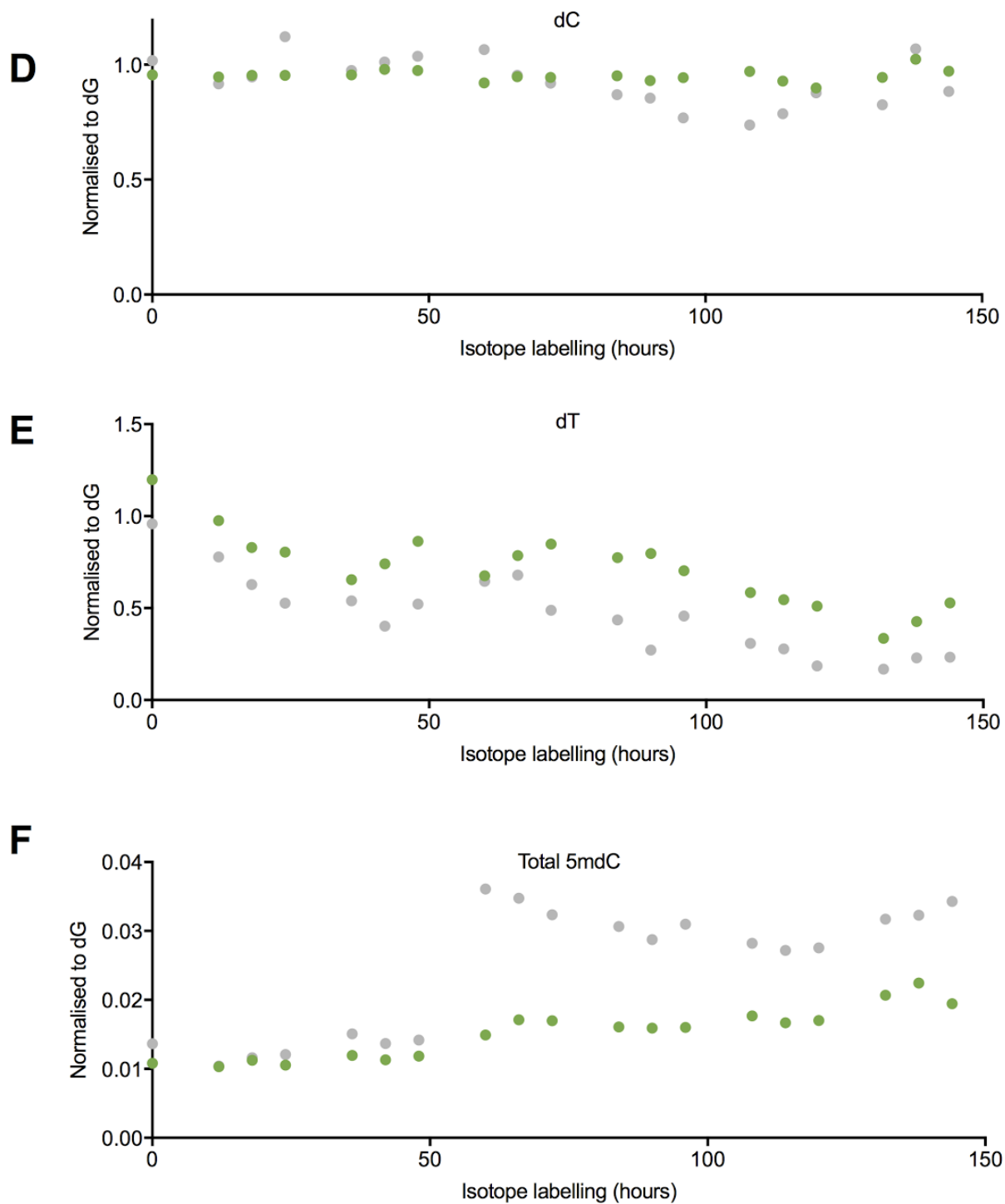


Figure 4.6 Stable isotope labelling of *Tet1* & *Tet2* double KO (DKO) mESCs

Tet1/2 double KO (DKO) mESCs and *Tet* Ctrl (WT) mESCs were maintained in FCS medium supplemented with CHIR99021 and PD0325901 for 5 passages. 12 hours after passaging, the medium was replaced with FCS + 2i *without* L-methionine. The medium was supplemented with ^{13}C -d₃ L-methionine and $^{15}\text{N}_2$ -uridine. Cells were cultured for 2 passages (144 hours). Cells were collected every 12 or 6 hours and the abundance of (a) C-Me (5mdC), (b) C*-Me*, (c) C-Me*, (d) dC, and (e) dT was quantified by LC-MS. The total 5mdC is also shown (f).

4.4. Discussion

Whether a result of aberrant *de novo* methyltransferase activity or potentially on-going active DNA demethylation, DNA methylation states appear to be more dynamic than previously thought. In this chapter, I have presented a LC-MS method to quantify dynamic changes in ‘steady state’ DNA methylation. Our preliminary *in vivo* results suggest that there is a degree of 5mdC turnover in post-mitotic cortical neurons. These results must be validated by BrdU labelling, to determine whether the observed incorporation of isotopomers is due to low levels of cell proliferation, active DNA repair, or cycles of DNA demethylation followed by *de novo* DNA methylation. Most studies of DNA methylation dynamics in neurons have looked at specific loci, where the addition and removal of cytosine methylation correlates with neuronal activity and the binding of transcription factors (Martinowich et al., 2003)(Murgatroyd et al., 2009). However, there are no published studies quantifying global 5mdC turnover. Rates of steady state DNA methylation turnover in the brain are thought to be higher than that of many cell types, and potentially play an important role in memory formation, neuronal plasticity and long term neuronal potentiation. It is unknown whether these findings can be reproduced in other cell types. This is in stark contrast to the long standing paradigm that cytosine methylation is a stable epigenetic mark.

Although *in vivo* models have great physiological relevance, stable isotope labelling *in vivo* is highly inefficient. In parallel to the experiment described in this chapter, my colleague Dr. R. Amouroux quantified stable isotope labelling during spermatogenesis (Amouroux et al., 2016). The entire duration of spermatogenesis takes approximately 34 days in mice. However, even after 10 weeks on a diet devoid of ‘light’ methionine, approximately 30% of 5mdC remained unlabelled. Similar observations have been made independently using different tissues. Therefore, even in proliferative cells, the incorporation of isotope labelled methyl groups is incomplete. Methionine is an essential amino acid, and therefore cannot be synthesised *de novo* by mammalian cells. Unlabelled methionine may be derived from impurities in the feed, such as additional proteins, or from endogenous recycled proteins. These observations present an additional hurdle for the study of

non-proliferative cells, where the proportion of labelled 5mdC is anticipated to be much lower than that seen in dividing cells. In this scenario, *ex vivo* cell culture-based approaches have the advantage of being highly simplified experimental models which can be easily manipulated.

Quantifying DNA methylation dynamics in mESCs presents a slightly different challenge, since most isotopomers will be incorporated via replication-linked maintenance DNA methylation. In this chapter, I have presented some of the initial experiments needed to optimise this experimental system. Notably, we found that the incorporation of ^{13}C -d₃ 5mdC was far more efficient than the incorporation of ^{15}N labelled dTTP and dCTP via the pyrimidine salvage pathway. Low levels of $^{15}\text{N}_2$ -labelled isotopomers are likely to reflect either the reliance on both the *de novo* and salvage pathways of pyrimidine biosynthesis or an abundance of exogenous or endogenous unlabelled uridine. FCS is an obvious source of unlabelled uridine. All mESCs were therefore cultured with dialysed FCS to reduce the pool of unlabelled uridine. Dialysed FCS (Biowest) uses a molecular weight cut-off (MWCO) of 10,000 Daltons. Therefore, short polynucleotides or uridine monomers would not be lost during dialysis. Any residual unlabelled uridine will compete with $^{15}\text{N}_2$ -labelled uridine. Typically, highly proliferative cells, including mESCs depend on *de novo* pyrimidine synthesis. Supplementing the culture medium with uridine should increase the cells reliance on the more energy efficient pyrimidine salvage pathway. Following the incubation of mESCs with either 75 μM and 100 μM $^{15}\text{N}_2$ -labelled uridine, the rates of pyrimidine isotopomer incorporation are similar; suggesting that the salvage pathway of pyrimidine synthesis is running at capacity. The metabolic switch from the *de novo* to the salvage biosynthetic pathway can be confirmed by labelling pyrimidines via the *de novo* pyrimidine pathway using, for instance, labelled glutamine.

Potentially, these challenges could be overcome by quantifying the rate of isotopomer *loss* following the transition from isotope-labelled medium to unlabelled medium. In this instance, the rate of isotopomer loss would depend on the rate of cell proliferation, and DNA methylation turnover but not the efficiency of isotopomer incorporation via the salvage biosynthetic pathway. Therefore, the downstream analysis should be simplified.

Nonetheless, mESCs are a potentially useful system for the study of dynamic, 'steady state' DNA methylation. We began our studies by looking at the role of TET proteins in 5mdC turnover. Although we see little differences in the rate of isotopomer incorporation, it will be worth repeating these experiments under the conditions described above. There are also numerous other proteins proposed to partake in the DNA methylation turnover. Genetic knockout (or knockdown) of these proteins could be used to assess the relative importance of each pathway in maintaining steady-state DNA methylation in mESCs. Alternatively, genetic ablation of DNA methyltransferases, individually and in combination, could highlight which enzymes are particularly important in the deposition of these marks.

Metabolic flux measures the production and consumption of metabolites in complex biological systems. Isotope tracing may also be useful for studying the flux of nucleoside isotopomers through the TET-mediated oxidation pathway. TET-mediated hydroxylation and BER DNA repair is widely accepted as a mechanism for the active removal of 5mdC. Logic dictates that TET-catalysed 5mdC hydroxylation should coincide with an increase of 5hmdC. However, this would not be the case if the oxidation of 5mdC to 5hmdC was rate limiting, and the subsequent oxidation of 5hmdC to 5fC and 5CaC was rapid. Quantification of 5mdC, 5hmdC, 5fC and 5CaC isotopomers should enable the rate of successive oxidation steps to be assessed. Assuming that TET-catalysed 5mdC hydroxylation and TDG-mediated excision is widespread in mESCs, these studies may enable us to identify the primary TDG substrate.

At specific loci, cyclic DNA methylation correlates with transcriptional activity (Metivier et al., 2008). Whether this occurs on a genome-wide scale is unknown. The dynamic relationship between DNA methylation and transcriptional activity can be assessed following the inhibition of productive elongation or genetic knockdown of putative elongation factors. Collectively my results point towards the idea that DNA methylation is in a constant state of flux. This idea is new and potentially controversial. If true, this would raise additional questions as to how the DNA methylome is safeguarded to prevent the accumulation of epimutations overtime; particularly in long-lived cells such as neurons.

Chapter 5

Regulation of global DNA methylation in mESCs
upon transition to 2i media

5.1 Introduction

Mouse embryonic stem cells (mESCs), derived from the inner cell mass (ICM) of the blastocyst, have traditionally been maintained in the presence of leukaemia inhibitory factor (LIF) and FCS. Under these conventional culture conditions, mESCs receive pro-differentiation signals but are constrained from differentiating by LIF (Evans and Kaufman et al., 1981)(Stewart et al., 1992). These ESC conditions harbour subpopulations of cells with distinct functional potentials and heterogeneous gene expression patterns (Guo et al., 2016). Alternatively, mESCs can be cultured in defined medium with two small molecule extracellular signal-regulated kinase (ERK) and glycogen synthase kinase 3 beta (GSK3 β) inhibitors (2i)(Ying et al., 2008). These inhibitors shield mESCs from differentiation-inducing signals whilst simultaneously promoting the expression of pluripotency-related genes. 2i conditions are thought to generate a homogeneous transcriptional state which is comparable to the E4.5 epiblast (Boroviak et al, 2014). Therefore dual ERK and GSK3 β is thought to block mESCs at a point prior to commitment and restriction of cell fate decisions.

In the genome of mESCs grown in FCS-containing medium, approximately 4% of all cytosines are methylated (Leitch et al., 2013). This is most prominent for CpG dinucleotides, which contain an average cytosine methylation level of approximately 80%. mESCs grown in 2i medium however are globally hypomethylated with approximately 2% of all cytosines reported to be methylated in 2i. Although it is often assumed that global DNA hypomethylation and the naive pluripotent state are functionally linked, this is yet to be proven. Importantly these states are reversible; mESCs can be ‘adapted’ to either condition when cultured for an extended period (up to 5 passages)(Leitch et al., 2013).

In 2013, three key papers were published that measured changes in DNA methylation in mESCs transferred from FCS-containing to 2i culture conditions (Leitch et al., 2013)(Ficz et al., 2013)(Habibi et al., 2013). The key conclusions drawn from these papers were as follows:

1. Upon transition from FCS-containing medium to 2i conditions, mESCs show substantial DNA demethylation in most genomic contexts including gene bodies, non-CpG island promoters that are substantially methylated in mESCs cultured in FCS conditions and also in the LINE and SINE transposon families.
2. IAPs and imprinted control regions (ICRs) are resistant to 2i-induced DNA methylation erasure.
3. Substantial downregulation of the *de novo* DNA methyltransferases accompanies the transition to 2i medium.
4. Expression of *Dnmt1* and its targeting factor *Uhrf1* does not change upon transition to 2i medium.
5. Cells grown in 2i medium upregulate *Prdm14*, a key factor that is responsible for the transcriptional repression of *de novo* DNA methyltransferase.

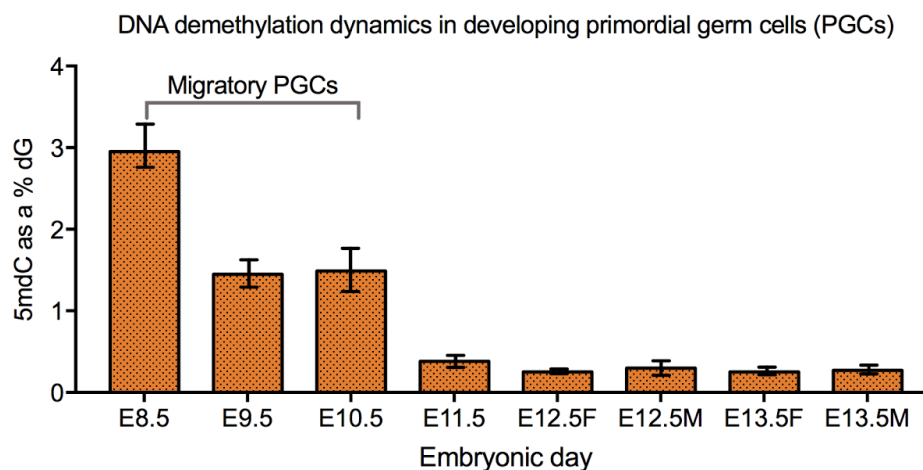


Figure 5.1 The dynamics of DNA demethylation in migratory PGCs is similar to that induced in mESCs upon transition to 2i media

Primordial germ cells (PGCs) were isolated from embryos between embryonic day (E) 8.5 and 13.5. From E12.5 onwards, male and female gonads were distinguished based on morphological differences. Isolated genomic DNA was digested and the abundance of 5mdC quantified by LC-MS. The analysis shows that DNA demethylation in PGCs occurs in two phases in (migratory and gonadal) respectively.

PGCs were isolated by Dr. P. Hill and LC-MS analysis was performed by Dr. R. Amouroux.

Of note, these changes are similar to the key events believed to drive genome-wide DNA demethylation in developing primordial germ cells (PGCs)(refer to Chapter 8 for a description of the DNA methylation dynamics

in PGCs). Therefore, 2i-induced reprogramming of mESCs to the pluripotent ground state is potentially a simple and versatile *in vitro* system to model the large scale reprogramming events in early migratory-stage PGCs. In particular, our analysis of 5mdC abundance in PGCs (measured and quantified by LC-MS using a method optimised for small cell numbers) clearly shows two phases of DNA demethylation (personal communication with Dr R. Amouroux). In migratory PGCs, 5mdC is approximately halved, after which a more extensive wave of DNA demethylation occurs (Fig. 5.1). Notably, both phases of DNA demethylation occur in the presence of DNMT1 (Leitch et al., 2013)(Hajkova et al., 2002). Ultimately, this system could be used to study how DNA methylation is lost despite the presence of DNMT1 activity.

5.2 Experimental aims

There are significant inconsistencies in the published data regarding the timing and mechanism of 2i-induced global DNA demethylation. This warranted further investigation into the timing and underlying mechanism of DNA demethylation and epigenetic reprogramming in mESCs upon the transition to 2i culture conditions. These issues may be addressed by studying the events immediately after transition to 2i. Specifically, I wanted to define the kinetics of DNA demethylation following the transition to 2i and explore the possible mechanisms behind 2i-induced *de novo* Dnmt downregulation.

5.3 Results

5.3.1. Experimental system

We decided to use a simple experimental system, whereby global levels of DNA modifications and the expression of epigenetic modifiers are quantified following the transition to 2i culture conditions. The results described in this chapter are based on five different cell lines (Fig. 5.1). Three were derived in FCS-based conditions whilst two (12g4 and 12g3) were derived in 2i. Previous studies have shown that epigenetic state, and DNA methylation levels in particular, reflect the current culture conditions rather than the derivation conditions (Leitch et al., 2013). Therefore, the 12g4 and 12g3 cell lines were adapted to FCS/Lif culture conditions for 5 passages prior to the transition back to 2i. To create a detailed picture of the events leading up to naive ground state pluripotency, mESCs were collected every 12 hours for 1 passage (72 hours) and after 5 passages in 2i; enabling the characterisation of both the immediate and steady-state response to 2i medium.

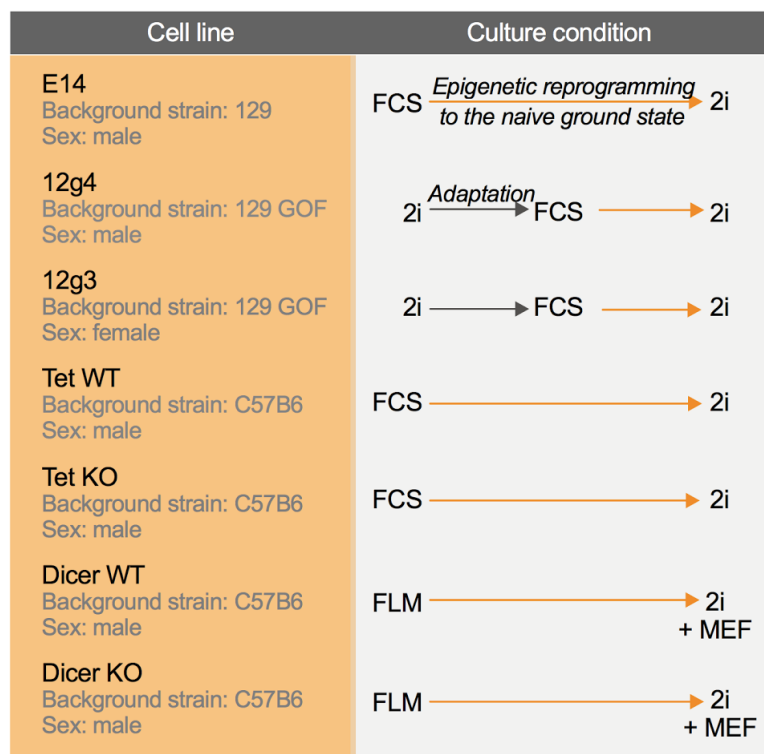


Figure 5.2 Cell lines and culture conditions used in this study

Abbreviations: FCS (Fetal calf serum-containing medium), 2i (serum-free medium with dual ERK and GSK3 β inhibition), FLM (FCS-containing medium on feeder cells), MEF (mouse embryonic fibroblasts), GOF (gain-of-function, $\Delta PE-Oct4$ GFP)

5.3.2. Dynamics of DNA modifications upon transition of mESCs to 2i culture conditions

LC-MS was used to accurately and simultaneously quantify global 5mdC and 5hmdC levels following the transition of mESCs from FCS-culture conditions to 2i medium. Figure 5.3 illustrates how different variables may affect the kinetics of 5mdC loss. Three biological replicates were analysed for each mESC line.

Currently, little is known about how genetic background affects DNA demethylation. Therefore, we decided to compare the kinetics of 2i-induced 5mdC loss in C57B6 and 129/ola mESCs. We found that C57B6 (*Tet* WT) mESCs cultured in FCS have significantly higher levels of global 5mdC compared to 129/Ola (E14) wild type mESCs (Fig. 5.3a). In C57B6 mESCs, the first 12 hours following the transition to 2i characteristically show a

large reduction in global 5mdC. In comparison, 5mdC loss in 129/Ola mESCs is more gradual. This may possibly reflect a larger degree of selection upon transition to 2i.

mESCs maintained in FCS/LIF culture conditions are considerably more heterogeneous than mESCs cultured in the presence of mouse embryonic fibroblasts (FCS/LIF/MEF, FLM). Thus we hypothesised that maintaining mESCs initially in FLM conditions would create a more synchronous, uniform phase of epigenetic reprogramming. In this scenario, the variation in global 5mdC or gene expression within the initial cell population may be reduced, potentially aiding a more reliable interpretation of the events associated with DNA demethylation. However, during the first 72 hours after transition to 2i, mESCs initially cultured in FLM conditions appear to be significantly hypermethylated; thus suggesting that not all MEFs were eliminated during the panning process.

Female (XX) mESCs are known to be hypomethylated compared to their male (XY) equivalents (Zvetkova et al., 2005), leading to suggestions that the expression of X-linked genes from both active X-chromosomes in XX mESCs may contribute to the reduced levels of 5mdC (Dr S. Ooi, personal communication). However, perhaps due to perceived difficulties with the derivation of XX mESCs in conventional FCS conditions, most studies of 2i-induced DNA demethylation have used XY mESCs. Therefore, little is known about the dynamics of DNA demethylation in XX mESCs upon transition to 2i. To assess whether 2i-induced changes in 5mdC are subject to sex-specific variation, XX (12g3) mESCs were transitioned to 2i and the global changes in DNA modifications analysed by the LC-MS described in Chapter 3. Importantly, 12g3 (XX) and 12g4 (XY) mESCs were derived from the same inbred mouse strain. Karyotyping, performed by Dr. M. Borkowska, determined that approximately 50% 12g3 mESCs have a distinct XO chromosome composition. In line with previous studies, 12g3 (XX/XO) mESCs maintained in FCS-culture conditions had significantly lower 5mdC compared to background matched 12g4 (XY) mESCs (unpaired t-test, * denotes a two-tailed p-value of <0.05)(Fig. 5.3c). However, the kinetics of 5mdC loss, and the global level of 5mdC in 2i are not significantly different between 12g4 (XY) and 12g3 (XX/XO) mESCs.

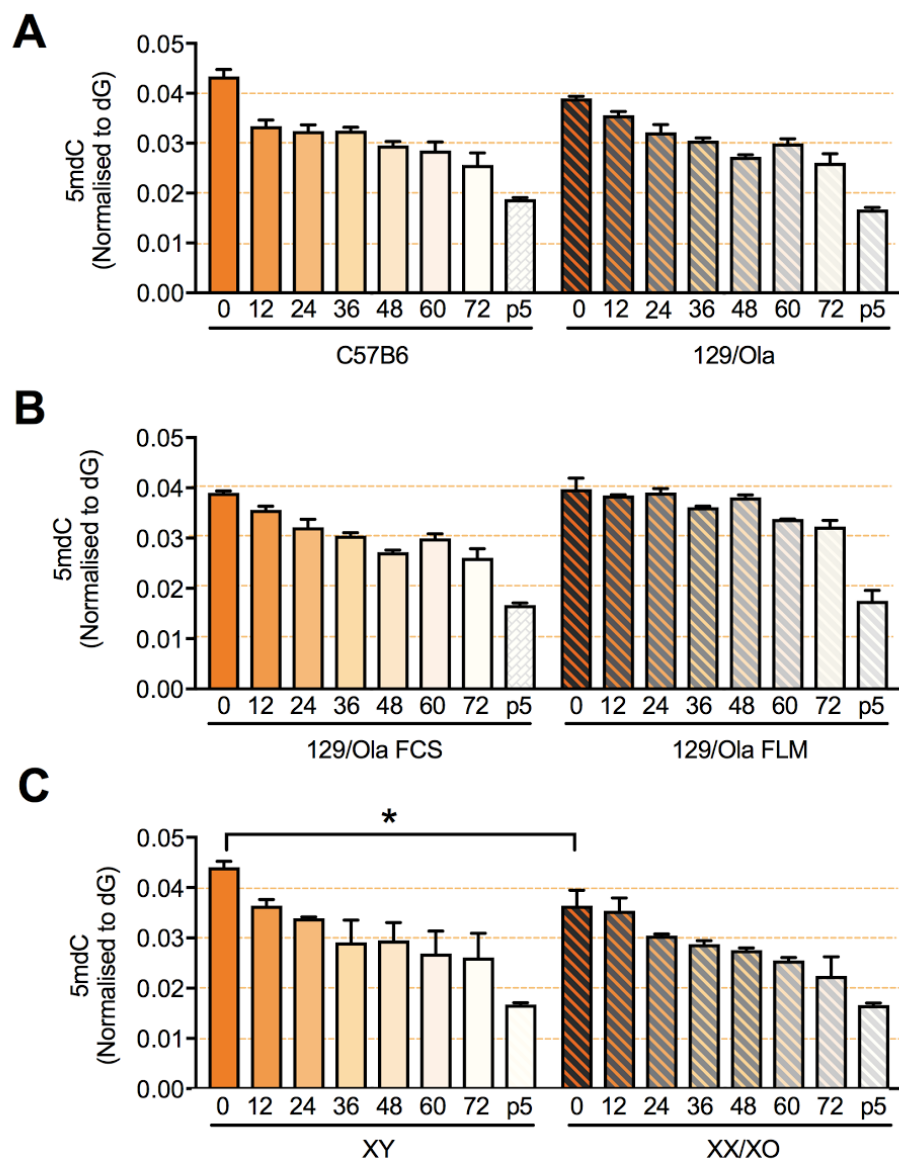


Figure 5.3 LC-MS quantification of 5mdC in mESCs upon transition to 2i

mESCs were transitioned from FCS-based growth medium to 2i medium. Cells were collected every 12 hours for 1 passage (72 hours) to assess the dynamics of 5mdC erasure. mESCs were further maintained in 2i medium for 5 passages to quantify the typically steady state 5mdC level. Extracted DNA was digested to nucleosides and the abundance of 5mdC was quantified by LC-MS. The results above have all been normalised to dG. Three independent biological replicates were collected for each sample. Error bars represent the standard error of the mean.

(a) Comparison of 5mdC loss in mESCs with different genetic backgrounds.

(b) Comparison of 5mdC loss in mESCs initially maintained in FCS or FLM culture conditions.

(c) Comparison of 5mdC loss in XY and XX/XO mESCs. Assuming a Gaussian distribution, the level of 5mdC in XY and XX/XO mESCs cultured in FCS conditions is significantly different (unpaired t-test, * denotes a 2-tailed p-value <0.05).

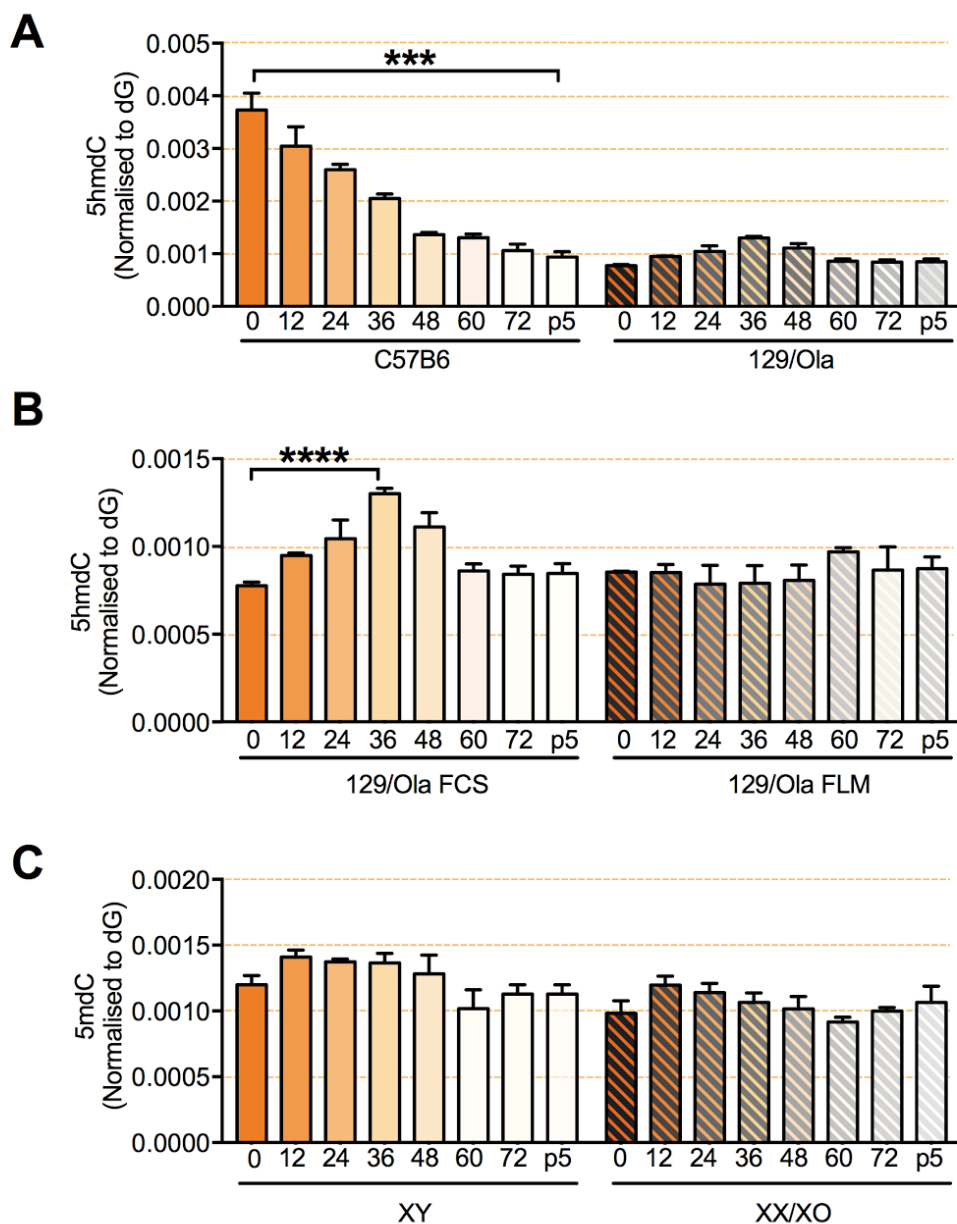


Figure 5.4 LC-MS quantification of 5hmdC in mESCs upon transition to 2i

(a) Comparison of 5hmdC dynamics between different genetic backgrounds (C57B6 & 129/Ola). 3 biological replicates were analysed. The chart shows the mean 5hmdC/dG and error bars represent the SEM. Extensive loss of 5hmdC is seen upon transition to 2i. This difference is highly significant (Unpaired t-test, *** denotes a p-value of 0.0001).

(b) Comparison of 5hmdC dynamics when mESCs are transitioned from feeder-free or feeder-dependent culture conditions to 2i. The increase in 5hmdC between 0 hours and 36 hours in 129.Ola mESCs initially cultured in FCS, feeder-free conditions is highly significant (Unpaired t-test, **** denotes a two-tailed p-value <0.0001).

(c) Comparison of 5hmdC dynamics between different male (XY) and female (XX/XO) mESCs upon transition to 2i.

These variables appeared to have a far greater affect on the level of 5hmdC. On the one hand, C57B6 mESCs have very high levels of 5hmdC in FCS culture conditions. Following the transition to 2i, 5hmdC appears to be diluted (fig. 5.4a). This is in stark contrast to the global 5hmdC in 129/Ola mESCs, which peak 36 hours after the transition to 2i before returning to its initial level (Fig. 5.4a and 5.4b). Meanwhile, in 12g3 and 12g4 mESCs, the global abundance of 5hmdC following the transition to 2i is unchanged (Fig. 5.4c). This observation may reflect the transient nature of 5hmdC generated by any 2i-induced TET activity.

The 12g4 mESC line has been used extensively by our group for investigating the effect of 2i-induced epigenetic changes (Leitch et al., 2013). To ensure that the work presented in this chapter is comparable with our previous studies, unless otherwise stated, the 12g4 mESC line was used in the experiments presented throughout this chapter.

5.3.3 Transcriptional downregulation of *de novo* Dnmts drive the rapid onset of DNA demethylation

DNA methylation levels are established and maintained by DNMT1, DNMT3A, and DNMT3B. At the transcriptional level, the *de novo* DNMTs (the mESC-specific *Dnmt3a* isoform 2, *Dnmt3b* and *Dnmt3l*) are downregulated within the first 12 hours after transition to 2i (Fig. 5.5a). The ubiquitous *Dnmt3a* isoform 1 continues to be expressed at a uniform, low level after transition to 2i. The downregulation of *Dnmt3a* isoform 2 and *Dnmt3b* was confirmed by western blot (Fig. 5.5b). We have found that expression of lamin B, unlike many other housekeeping genes, does not change upon transition to 2i (personal communication with Dr K. McEwen). Therefore lamin B was chosen as a loading control. Fig. 5.5c shows the normalised signal intensity for *Dnmt3a* isoform 2 and *Dnmt3b* isoform 1, quantified using ImageJ.

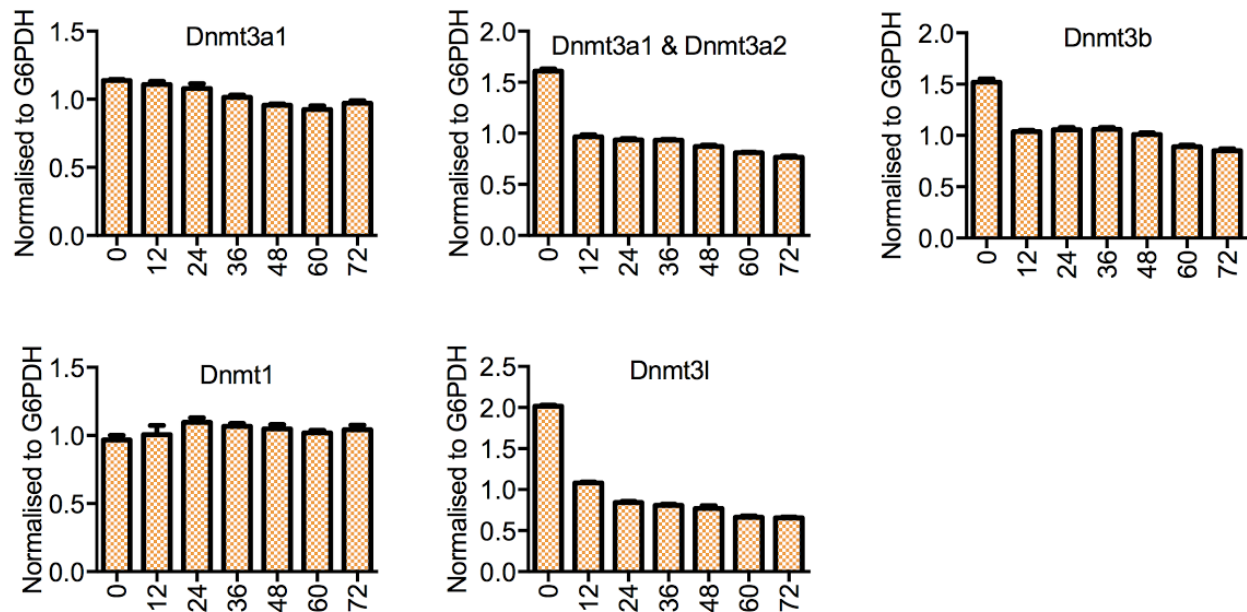
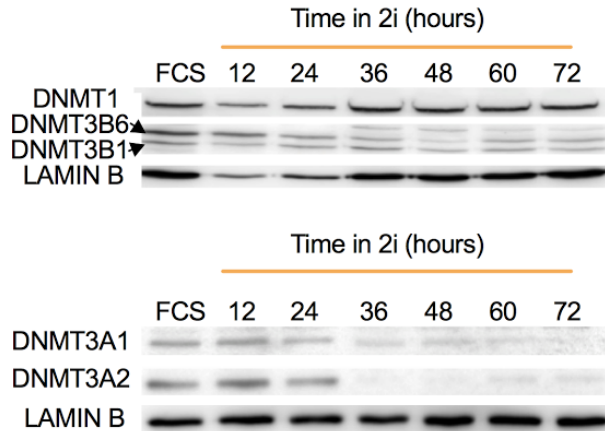
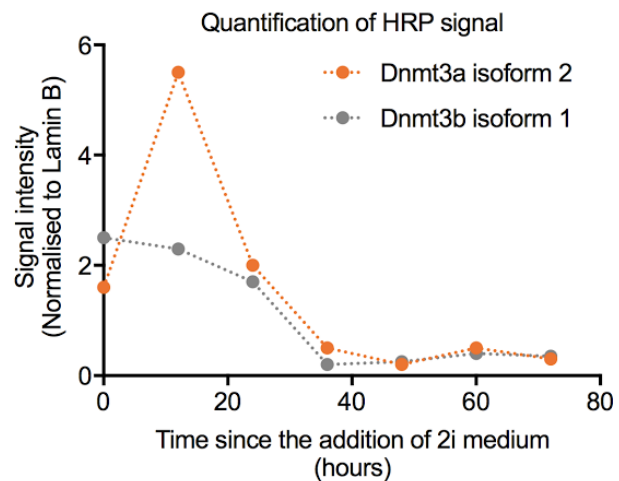
A**B****C**

Figure 5.5 Transcriptional downregulation of *de novo* Dnmts occurs rapidly upon transition of mESCs to 2i

(a) RT-qPCR quantification of Dnmt transcripts following the transition of mESCs to 2i culture conditions. Gene expression was quantified using SYBR green and the standard curve method. It has been determined that G6PDH is expressed uniformly in mESCs cultured in FCS and 2i culture conditions. Therefore, all inputs are normalised to G6PDH.

(b) Western blot of DNMT expression in mESCs following the transition to 2i. Lamin B is expressed uniformly in mESCs cultured in FCS and 2i conditions and was therefore chosen as a suitable loading control.

(c) Quantification of HRP signal intensity using image J. All signals were normalised to lamin B and the background signal.

Dnmt1 gene and protein expression is uniform in both conventional FCS and 2i culture conditions (Fig. 5.5a and 5.5b), suggesting that 2i-induced global DNA demethylation is linked to decreased expression (or post-transcriptional regulation) of the *de novo* Dnmt rather than impaired DNMT1 protein stability. This data shows that, DNMT1 alone is unable to efficiently compensate for the loss of *de novo* DnmTs. Any potential contribution of DNMT1 to the hypomethylated phenotype must therefore be on the level of sequestration or failed recruitment.

5.3.4. Active and passive mechanism contribute to DNA demethylation

Both active and passive mechanisms can contribute to DNA demethylation. Despite their designation as *de novo* DNA methyltransferases, DNMT3A and DNMT3B are also responsible for the stable inheritance of DNA methylation patterns. DNMT3L stimulates DNMT3A and DNMT3B. Therefore, significantly reduced *Dnmt3l* expression might also impact on maintenance methylation in 2i by impeding residual DNMT3A and DNMT3B activity. To test this hypothesis, it may be useful to constitutively express *Dnmt3l* in mESCs and compare the kinetics of DNA demethylation; with any significant reduction in demethylation indicating that loss of DNMT3L-mediated DNMT3A and DNMT3B activity contributes to 2i-induced DNA demethylation.

Since passive DNA demethylation is a replication-dependent process, changes in the rate of cell proliferation may affect the rate of 5mdC dilution. To assess whether cell cycle dynamics differ between mESCs cultured in FCS and 2i conditions, 12g4 mESCs 'adapted' to either condition were incubated with EdU for 15 minutes and dual labelled with an Alexa fluor azide and DAPI. In 2i culture conditions, the proportion of cells residing in S-phase (66%) was lower than the proportion of S-phase mESCs (73%) in FCS-culture conditions (Fig. 5.6a and 5.6b). Furthermore, the rate of mESC proliferation, quantified using an MTT assay and cell counting, does not differ substantially between FCS and 2i culture conditions (Fig. 5.6d). Thus confirming that mESCs cultured in 2i-conditions have a shorter S-phase compared to mESCs cultured in FCS-conditions. Therefore, it is possible that the dynamics of DNA synthesis, pre-replication complex assembly or recruitment of DNMT1 to the replication foci may differ depending on the culture condition.

Typically, DNMT1 is responsible for maintaining DNA methylation states through cell divisions. Anything preventing faithful recruitment of DNMT1 to replication foci could potentially contribute to passive DNA demethylation. UHRF1 is a DNMT1 interacting factor and is involved in recruiting maintenance methyltransferase to the replication foci. We quantified the expression of UHRF1 in mESCs cultured in FCS conditions and following the transition to 2i (Fig. 5.6e). UHRF1 however appears to increase upon transition to 2i, suggesting that failed maintenance methylation is not the main driver of 2i-induced DNA demethylation.

Alternatively, DNA demethylation may be triggered by TET-mediated 5mdC hydroxylation. Given that hmdC is an oxidation product of 5mdC and an intermediate in TET-driven DNA demethylation, elevated hmdC may be indicative of increased TET-activity. *Tet1* and *Tet2* are highly expressed in mESCs (Ito et al., 2010). At the protein level, TET1 and TET2 appear to be upregulated following the transition to 2i (Fig. 5.7d). TET expression in 'steady state' 2i conditions however, is reduced. Indicating that potentially TET activity coincides with DNA demethylation. Therefore, I wanted to investigate the effect, if any, of *Tet1* and *Tet2* deletion on the kinetics of epigenetic reprogramming to the pluripotent ground state. *Tet1* and *Tet2* double knockout (DKO) mESCs (predominantly C57B6 genetic background) were given to us by Prof. R. Jaenisch (MIT). Following the transition to 2i, LC-MS quantification of 5mdC illustrates that the kinetics of DNA demethylation in *Tet* DKO cells is notably slower than that of the control wild type mESCs. After 72 hours in 2i *Tet* DKO mESCs have significantly higher 5mdC levels than *Tet* WT mESCs (calculated using an unpaired t-test, whereby *** denotes a p-value < 0.001) (Fig 5.7a). However, substantial DNA demethylation does nonetheless occur in *Tet* DKO cells, suggesting that TET activity is not essential for DNA demethylation and thus 2i-induced bulk DNA demethylation is not a TET-dependent process.

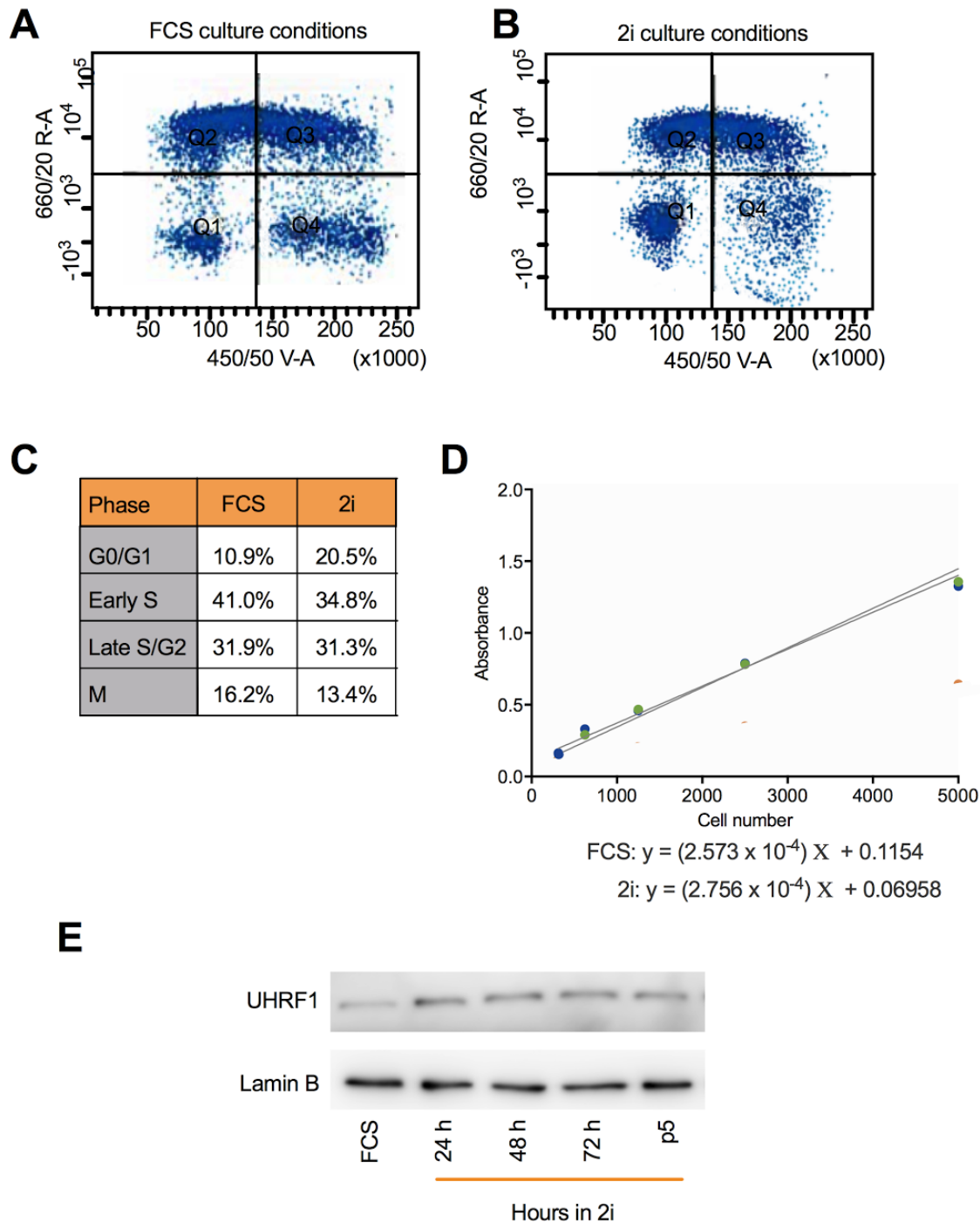


Figure 5.6 Culture-induced changes to the cell cycle and proliferation dynamics

mESCs adapted to either (a) FCS or (b) 2i culture conditions were labelled with EdU for 15 minutes. The proportion of mESC in G0/G1, Early-S, Late G/G2, and M-phase was visualised by further DAPI staining and flow cytometry. (c) Calculated percentages of cells in each cell cycle phase. (d) A colorimetric MTT assay was performed to assess how the rate of proliferation differs between mESCs cultured in FCS and 2i-culture conditions. Serial dilutions of mESCs were plated in triplicate and allowed to proliferate for 12 hours before the start of the assay. (e) 12g4 mESCs maintained in FCS culture conditions were transferred to 2i condition. The expression of UHRF1 was visualised by western blot during the first 72 hours after the addition of 2i and after 5 passages in 2i. Lamin B was used as a loading control.

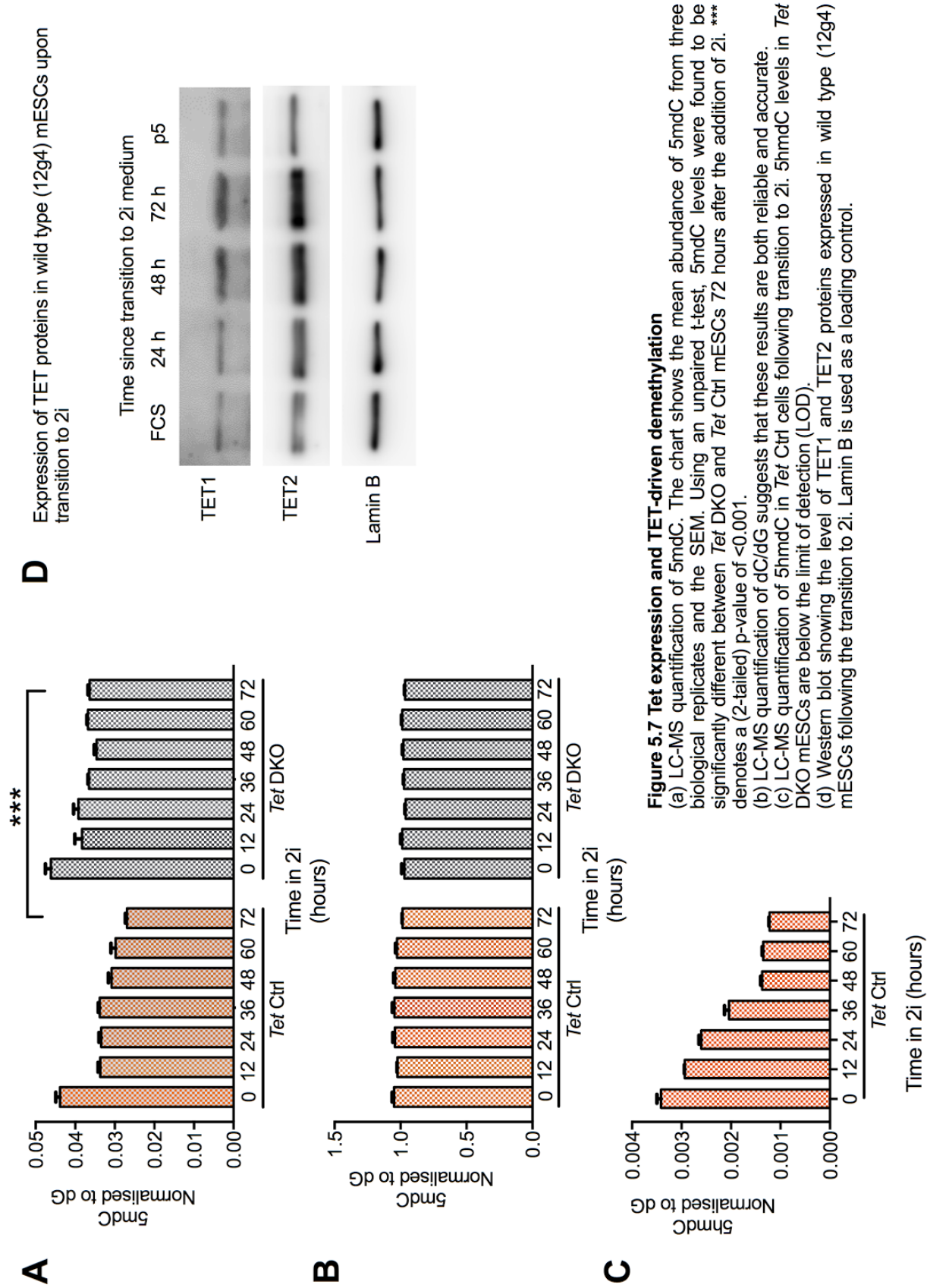


Figure 5.7 Tet expression and TET-driven demethylation

(a) LC-MS quantification of 5mdC. The chart shows the mean abundance of 5mdC from three biological replicates and the SEM. Using an unpaired t-test, 5mdC levels were found to be significantly different between *Tet* DKO and *Tet* Ctrl mESCs 72 hours after the addition of 2i. *** denotes a (2-tailed) p-value of <0.001.

(b) LC-MS quantification of dC/dG suggests that these results are both reliable and accurate.

(c) LC-MS quantification of 5hmdC in *Tet* Ctrl cells following transition to 2i. 5hmdC levels in *Tet* DKO mESCs are below the limit of detection (LOD).

(d) Western blot showing the level of TET1 and TET2 proteins expressed in wild type (12g4) mESCs following the transition to 2i. Lamin B is used as a loading control.

5.3.5. Dynamics of Non-CpG and CpG DNA demethylation

Non-CpG methylation is catalysed exclusively by DNMT3A and DNMT3B. Since the *de novo* Dnmts are specifically downregulated in 2i, I reasoned that the initial loss of 5mdC may include a disproportionately large loss of non-CpG methylation. To test this hypothesis, Dr. P. Hill used reduced representation bisulphite sequencing (RRBS) data (Accession GSE77420) to calculate the proportion of CpG, CpA and CpT methylation in E14 mESCs 24, 48, 72, 96 and 168 hours after transition to 2i conditions. However, we could not find any evidence that non-CpG dinucleotides were more likely to undergo DNA demethylation upon the transition to 2i medium.

5.3.6. Prdm14 transcriptional upregulation is not characteristic of early 2i-mediated epigenetic reprogramming

The mechanism driving *de novo* Dnmt repression in mESCs grown in 2i culture conditions is generally considered to involve the upregulation of the transcriptional repressor *Prdm14*. For instance, *Prdm14* null cells show high levels of 5mdC in both FCS-containing and 2i culture conditions (Leitch et al., 2013)(Fig. 5.8). However, quantification of *Prdm14* transcript levels after the initial transition to 2i suggests that this may not be the complete story (Fig. 5.8a). Although *Prdm14* is highly expressed in mESCs cultured in 2i conditions, the initial upregulation of *Prdm14* is modest; during the first 72 hours after transition to 2i, relative *Prdm14* expression is only a fraction of that seen in mESCs adapted to 2i (Fig. 5.8a)(Leitch et al., 2013). Nonetheless *Dnmt3a* and *Dnmt3b* are invariably repressed and global 5mdC levels reduced (Fig. 5.8a). Of course, we must confirm that PRDM14 protein expression is not increased. However, traditionally PRDM14 antibodies have suffered from poor specificity and therefore we have been unable to find a suitable antibody to quantify PRDM14 protein expression.

Based on these observations, I propose that *Prdm14* upregulation is not the initial trigger for *de novo* Dnmt repression. Nevertheless, since PRDM14 is known to be key link between the pluripotency and epigenetic transcriptional networks in mESCs, its presence is likely to be important for epigenetic reprogramming; perhaps being required for the maintenance of *de novo* Dnmt repression (Leitch et al., 2013).

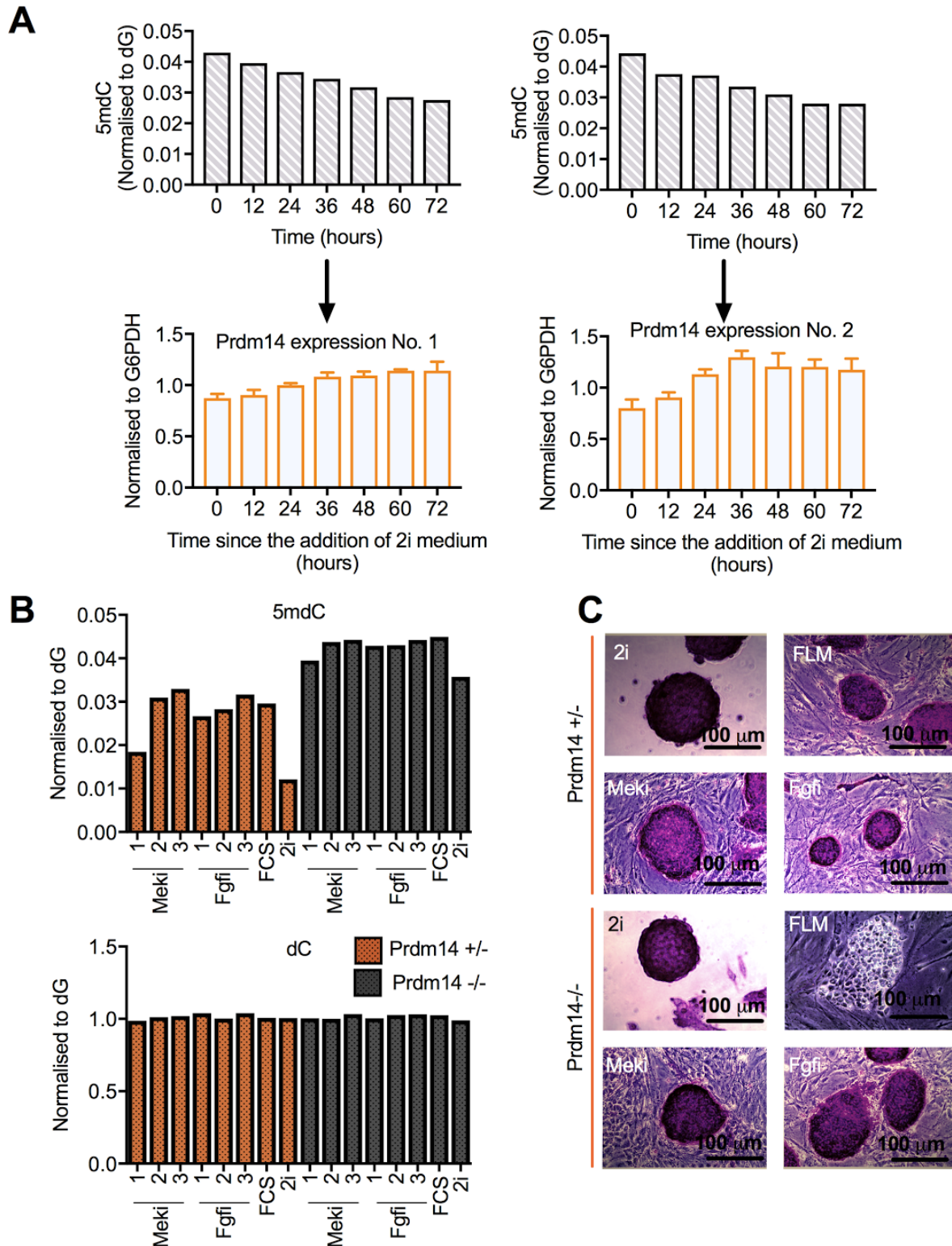


Figure 5.8 Prdm14 expression and DNA demethylation in 2i

(a) LC-MS quantification of 5mdC loss in wild type mESCs following transition to 2i. DNA demethylation proceeds rapidly. However the transcriptional upregulation of Prdm14, as quantified by RT-qPCR, is much slower.

(b) LC-MS quantification of 5mdC in Prdm14 +/- and Prdm14 -/- cells.

(c) Alkaline phosphatase staining of Prdm14 +/- and Prdm14 -/- cells maintained in FLM, FLM supplemented with either Mek inhibitor or Fgf inhibitor, and 2i.

5.3.7. *Prdm14* is not solely responsible for *Dnmt* downregulation

To assess the contribution of PRDM14 to the maintenance of the pluripotent ground state, we have obtained *Prdm14* null mESCs from Prof. A. Surani (University of Cambridge). Although *Prdm14* null cells proliferate and maintain an undifferentiated state in 2i, when cultured in conventional FLM conditions, they show decreased expression of pluripotency markers and alkaline phosphatase staining along with increased expression of lineage specific genes indicative of loss of pluripotency (Yamaji et al., 2013). As reported, this is due to the cells being highly sensitive to FGF signalling and thus easily primed for differentiation. *Prdm14* null cells can however be maintained in FLM conditions in the presence of either FGF- or ERK-inhibitors (Fig. 5.8c).

To confirm whether *Prdm14* upregulation is required for *de novo* *Dnmt* downregulation, I quantified the level of 5mdC in *Prdm14* null cells adapted to either 2i or FLM + FGF/ERK inhibitor culture conditions. The expression of the DNA methyltransferases was also quantified by RT-qPCR to ensure that the presence of FGF or ERK inhibitors did not alter their expression. *Prdm14* null mESCs are more highly methylated than *Prdm14* +/- mESCs of the same genetic background, regardless of the culture condition (Fig 5.8b) (see also Leitch et al., 2013). However, *Prdm14* null mESCs cultured in 2i conditions also show slightly reduced levels of 5mdC compared to *Prdm14* null mESCs. PRDM14-independent mechanisms of *de novo* DNA methyltransferase regulation must therefore exist. These mechanisms may contribute to the *de novo* DNA methyltransferase downregulation observed in wild type mESCs upon transition to 2i.

5.4 Discussion

We have found that DNA demethylation occurs rapidly upon transition to 2i, with approximately 25% of total 5mdC being lost in the first 72 hours. This process appears to be partially regulated at the transcriptional and post-transcriptional level.

Previously, it has been shown that loss of 5mdC in mESCs cultured in 2i growth conditions is not due to the loss of maintenance methyltransferase activity (Bostick et al, 2007)(Rothbart et al., 2012)(Sharif et al, 2007). Specifically, transcript levels of *Dnmt1*, *Uhrf1* and PCNA do not change between 2i and FCS culture conditions and *Dnmt1* protein levels and catalytic activity are unchanged in 2i. Additionally, 2i conditions do not alter the DNMT1/UHRF1/PCNA complex formation or its enzymatic activity (Habibi et al, 2012), suggesting that DNA demethylation is caused by reduced *de novo* methyltransferase activity and possibly the enzymatic loss of 5mdC.

More recently, this hypothesis has been revisited. Genetic ablation of either *Dnmt1* or *Uhrf1* prior to the transition to 2i increases the rate of DNA demethylation (Von Meyenn et al. 2016). This, is not surprising, since *Dnmt1* KO mESCs are known to be severely hypomethylated; approximately 20% 5mdC compared to wild-type cells. However, quantification of UHRF1 protein levels in mESCs cultured in FCS and 2i growth conditions indicates that expression of UHRF1 is reduced upon transition to 2i and therefore DNA demethylation may partially be due to reduced DNMT1 recruitment to replication foci. Additionally the expression of a UHRF1-GFP fusion protein, quantified by FACS, decreased rapidly upon transition to 2i. We however, did not observe a loss of UHRF1 at the protein level. Considering the inconsistencies presented to date, it may be worthwhile quantifying DNA demethylation dynamics in mESCs which constitutively express the *de novo* or maintenance Dnmts to assess their relative contributions to DNA methylation following transition to 2i culture conditions.

Initially, 2i-induced DNA demethylation was attributed, at least in part, to TET-driven hydroxylation of 5mdC (Ficz et al., 2013). Accordingly, I have found that in *Tet* DKO mESCs, 2i-induced DNA demethylation is slower, and less extensive than the demethylation seen in wild type counterparts. However, other recent studies have shown that the kinetics of 2i-induced DNA demethylation in *Tet* triple KO (TKO) mESCs are remarkably similar to that seen in wild type (E14) cells (Von Meyenn et al, 2016). Notably, in this study the *Tet* TKO cells were derived from a C57B6 background whilst E14 mESCs are from 129/ola mice. In our experience, the background of a cell line, and particularly C57B6 lines, can affect DNA methylation. Specifically we routinely quantify 5mdC in FCS-cultured C57B6 mESCs at above 4%; significantly higher than cells on a predominantly 129 background (E14 or 12G4). Therefore, care must be taken when comparing demethylation kinetics across different backgrounds. Surprisingly, in this study, the *Tet* TKO and background control mESCs have very low levels of 5mdC in FCS; Von Meyenn records approximately 2% for *Tet* wild type cells and 3% for *Tet* TKO cells. In my experience, DNA demethylation is consistently triggered rapidly upon the addition of 2i, however during the first 48 hours, no loss of DNA methylation is seen in the *Tet* wildtype cell line.

It has previously been shown that supplementation of mESC culture medium with vitamin C triggers extensive DNA demethylation by the upregulation of TET activity (Blaschke et al., 2013)(Young et al., 2015)(Chung et al., 2010) and enhances the generation of iPSCs from terminally differentiated cells, a process typically accompanied by global loss of 5mdC (Esteban et al., 2010)(Stadtfeld et al., 2012). Therefore, I believe that TET-triggered hydroxylation is *potentially* a driver of 2i-induced DNA demethylation. However, further work is needed to confirm this. Biochemical analysis indicates that the preferred substrate of TET proteins is 5mdC (Hu et al., 2015). Therefore, measuring TET activity based on 5hmdC seems a reasonable approach. However, our understanding of TET-triggered 5mdC hydroxylation is far from complete, and it is possible that different cofactors, cosubstrates or proteins associated with TET could affect the efficiency of this reaction. Therefore, we must bear in mind that a decrease in 5hmdC could be indicative of either reduced oxidation of 5mdC to 5hmdC or increased turnover of 5hmdC. Instead, the system described in the previous chapter could be used to quantify the flux of oxidised 5mdC derivatives from 5mdC to 5fC and 5CaC, and therefore indicate if the *rate* of TET-mediated oxidation is altered following the addition of 2i.

The transcriptional regulator Prdm14 is widely believed to be responsible for the downregulation of the *de novo* DNA methyltransferases upon transition to 2i. However, although hypermethylated compared Prdm14 +/- mESCs, Prdm14 -/- cells nonetheless have lower levels of 5mdC in 2i compared to FCS culture conditions. Additionally, the kinetics of Prdm14 transcriptional upregulation does not appear compatible with the kinetics of *de novo* Dnmt downregulation. Therefore, I suggest that Prdm14 may not initiate *de novo* Dnmt downregulation; instead being important for maintaining low level *de novo* Dnmt expression in ground state pluripotent cells. Alternatively, Prdm14 may indirectly be upregulated following the increased expression of the pluripotency transcriptional network.

Chapter 6

miRNAs as regulators of naive and ground state
pluripotency

6.1 Introduction

6.1.1 MicroRNA

miRNAs are endogenous, short non-coding RNAs that post-transcriptionally regulate gene expression. Regulation typically occurs by mRNA sequestration, mRNA degradation or translational inhibition (Wu et al., 2006). However there are also accounts that miRNAs can stabilise transcripts, prolonging their half life. The mode of post-transcriptional regulation is thought to depend on the complementarity between 2-8 nucleotides at the 5' end of the mature miRNA (the seed region) and the target mRNA (Farazi et al., 2008); with perfect complementarity causing transcript degradation whilst partial complementarity inhibits translation. miRNA regulation is widespread. Approximately one third of protein coding genes are thought to be regulated by miRNAs (Lewis et al., 2005). It is therefore unsurprising that miRNAs are involved in the regulation of diverse physiological and pathophysiological processes.

miRNAs can be located within various genomic contexts. Approximately half of all miRNAs are located within intergenic regions whilst others reside within the introns of protein coding genes; either being cotranscribed with the host gene or being transcribed using their own promoter. miRNA loci in close proximity can form polycistronic transcription units (clusters). Typically miRNAs within a single cluster are co-transcribed by Pol II, but can be post-transcriptionally regulated independently.

6.1.2. miRNA biogenesis

There are several endonucleolytic processing steps between the biogenesis of a primary miRNA (Pri-miRNA) and a mature miRNA. Pri-miRNAs are long (typically greater than 1 kb) hairpin structures (Lee et al., 2002). The nuclear RNase III DROSHA and its essential cofactor DGCR8 form a microprocessor which crops the

pri-miRNA stem loop; releasing a hairpin structure of approximately 65 nt (pre-miRNA)(Denli et al., 2004)(Borchert et al., 2006). Pre-miRNAs are exported by a nuclear pore complex into the cytoplasm where miRNA maturation continues (Yi et al., 2005). In mammals, DICER binds to both the 5' phosphorylated end of the pre-miRNA and 3' end of the dsRNA to determine the cleavage sites. This is known as the 5' counting rule and the 3' counting rule respectively (MacRae et al., 2007)(Park et al., 2011). The remaining RNA duplex is loaded onto AGO proteins to form an RNA-induced silencing complex (RISC). Subsequent removal of the passenger strand produces a mature RISC; consisting of an Ago protein and the guide strand.

Non-canonical pathways of miRNA biogenesis, described based on the deep sequencing of cells deficient in DROSHA, DGCR8 and DICER, suggest that Microprocessor and DICER-independent pathways must exist (Cheloufi et al., 2010)(Ruby et al., 2007)(Okamura et al., 2007). Approximately 1% of conserved miRNAs are produced by these non-canonical pathways; including miR-320 and miR-451 families (Xie et al., 2013). The vast majority of these alternative pathways bypass DROSHA-mediated processing, for example by mRNA splicing or by the direct transcription of short hairpin RNAs.

6.1.3. miRNA and pluripotency

miRNAs are believed to be important regulators of 'stemness' in mESCs; specifically being associated with proliferation, differentiation and reprogramming (Bartel et al., 2004)(Houbaviy et al., 2003). Many miRNAs are exclusively expressed in pluripotent cells; with some being specifically associated with either primed or naive states (Jouneau et al., 2011). Others are expressed at only low levels following differentiation.

Studies using both *Dicer* ^{-/-} and *Dgcr8* ^{-/-} mESCs have furthered our understanding of miRNAs and pluripotency (Kanellopoulou et al., 2005)(Murchison et al., 2005)(Wang et al., 2007)(Kumar et al., 2015). Deletion of either gene is embryonic lethal (Bernstein et al., 2005). However, the ability to derive homozygous knockout mESCs demonstrates that neither are necessary for pluripotency *per se*. Nevertheless, phenotypically, *Dicer* ^{-/-} and *Dgcr8* ^{-/-} mESCs have severe defects relating to both proliferation and development. For example,

loss of *Dicer* increases the proportion of cells in G0/G1 (Murchison, 2005), suggesting that in wild type mESCs, miRNAs are necessary for the suppression of the restriction point and commitment to S-phase. Additionally, deletion of either microprocessor abrogates differentiation into embryoid bodies; differentiating *Dicer* $-/-$ cells lack markers of the endodermal and mesodermal lineages (Kanellopoulou, 2005) whilst in *Dgcr8* $-/-$ cells, the upregulation of differentiation markers is severely delayed following withdrawal of LIF (Wang, 2007). *Dgcr8* $-/-$ ESCs are also unable to generate highly differentiated teratomas following subcutaneous injection; indicating that both DICER and DGCR8 are required for the transition from pluripotency to lineage commitment (Wang et al., 2007).

6.3.1.1. The miR-290 cluster

A large portion of these mESC-specific miRNAs belong to just two polycistronic clusters; the miR-290 family and the miR-302 family (Mathieu et al., 2013)(Barrosa-del et al., 2009). The conserved (at least in placental mammals) miR-290 polycistronic cluster is the most abundant in mESCs; with its 14 mature miRNAs totaling 60% of the miRNA pool (Marson et al., 2008). Expression of the cluster is downregulated rapidly upon mESC differentiation and is obsolete in adult mouse organs. Prominent targets including p21, Rbl2, caspase 2 and Ei24, indicate that the miR-290 cluster is involved in regulating the G1/S transition, *de novo* DNA methylation and protecting cells from genotoxic stress (Wang, 2008)(Sinkkonen, 2008), (Benetti, 2008). Of these, Rbl2 is a known transcriptional repressor of *Dnmt3a*, *Dnmt3b* and *Dnmt3L* with two conserved miR-290 binding sites within its 3'UTR. siRNA directed against Rbl2 results in an upregulation of *de novo* Dnmts whilst overexpression of the miR-290 cluster downregulated all three *de novo* Dnmts; thus linking mESC-specific miRNA and DNA methylation (Sinkkonen et al., 2008). *Dnmt3a* and *Dnmt3b* gene expression has previously been found to be downregulated in *Dicer* KO cells (Benetti, 2008). During differentiation, methylation and subsequent silencing of the *Oct4* promoter is inefficient. Transfection of either *de novo* Dnmt or the entire miR-290 cluster rescues this phenotype, indicating that deficient DNA methylation contributes to the differentiation defects seen in *Dicer* KO mESCs.

6.3.1.2. The miR 302 cluster

The miR-302 cluster also controls cell proliferation; ectopic expression reduces the proportion of cells in G1 in both primary and transformed cells, whilst concomitantly increasing the proportion of cells in S-phase via inhibition of cyclin D1 translation (Wang et al., 2013). Additionally, miR-302b was found to indirectly regulate *Oct4* gene expression and targets cyclin D2, an important developmental regulator during gastrulation (Lee, 2008). Specific miRNAs are known to regulate naive and primed pluripotency. Whilst the miR-290 cluster is supposed to regulate the naive pluripotent state, the miR-302 cluster regulates primed pluripotency. Ultimately however, the precise role of miRNA in differentially regulating ground-state, naive and primed pluripotent stem cells is currently unknown.

6.1.4. miRNAs enhance the efficiency of reprogramming

More recently, mESC-specific miRNAs have been characterised as *bona fide* reprogramming factors. For instance, a handful of studies have claimed that miRNAs, including miR-291-3p, miR-294 and miR-295 (all belonging to the miR-290 cluster), are capable of inducing pluripotency in somatic cells in the absence of exogenous transcription factors (Anokye-Danso, 2011; Lin, 2008; Miyoshi, 2011)(Judson, 2009). Furthermore, fibroblasts deficient of all mature miRNA cannot be reprogrammed to iPSCs (Kim, 2012).

6.2 Experimental aims

Since undifferentiated *Dicer* KO mESCs can be maintained in FCS-containing medium, miRNAs are not considered to be important for the maintenance of pluripotency *per se*. However, many miRNAs are known to be specifically expressed in pluripotent cells, where they are presumably important regulators of gene expression. I wanted to interrogate the role of miRNAs in defining the epigenetic and transcriptional landscape of mESCs. In particular, very little is known about miRNA-mediated regulation of ground state pluripotency.

By sequencing the transcriptomes of *Dicer* KO mESC maintained in 2i and FCS-containing medium, I hoped to further our understanding of miRNA-mediated processes in the pluripotent ground state.

Dicer KO mESCs are known to be hypomethylated (Nesterova et al., 2008). Since lineage commitment and differentiation coincide with changes in global DNA methylation, it is perhaps not surprisingly that *Dicer* KO mESCs have severe developmental defects. I questioned whether *Dicer* KO cells would also be unable to erase 5mdC during 2i-induced reprogramming and therefore whether miRNAs may have a more general role in regulating epigenomic plasticity and large-scale reprogramming.

As shown in Chapter 5, downregulation of *de novo* Dnmts does not appear to be initiated by PRDM14. Since DNMT3A and DNMT3B are known to be indirectly regulated by miRNAs (Benetti et al., 2008), miRNAs may also regulate the 2i-induced downregulation of *de novo* Dnmts. Using small RNA-sequencing datasets produced by FCS- or 2i-cultured mESCs, I wanted to identify miRNAs specifically involved in 2i-induced reprogramming and particularly those that may directly or indirectly regulate *de novo* Dnmt expression.

6.3 Results

6.3.1. miRNA and the post-transcriptional regulation of *de novo* Dnmts in ground state ESCs

Previously, our group discovered that enzymes involved in miRNA processing are differentially regulated in mESCs cultured in FLM and 2i conditions (personal communication with Dr K. McEwen). Specifically, *Drosha* and *Dicer* are significantly transcriptionally downregulated in 2i conditions (Fig. 6.1) whilst *Ago2* and *Ago3* are upregulated. Therefore I asked whether miRNAs could regulate the transition to ground state pluripotency. To address this question, we have obtained *Dicer* KO mESCs from Prof. N. Brockdorff (Oxford). The cells were derived from *Dicer*^{-/-} blastocysts; their genetic background being predominantly C57B6. Regardless of the basal medium, maintenance of pluripotent *Dicer* KO cells is crucially dependent on co-culture with MEFs (Fig. 6.1b). Unlike most mESCs, the density of MEFs is critical, with *Dicer* KO cells only proliferating on a highly confluent feeder layer.

To determine whether miRNAs are involved in 2i-induced DNA demethylation and *de novo* Dnmt downregulation, *Dicer* KO cells were transitioned to 2i + MEF conditions. mESCs were collected every 24 hours for 5 days and separated from the feeder cells based on the expression of the pluripotency cell marker SSEA1 by FACS (Fig. 6.1c). The global abundance of 5mdC in each sample was quantified by LC-MS (Fig. 6.2). As previously shown, *Dicer* KO cells are hypomethylated in FCS-conditions compared to wild type mESCs. Interestingly, in *Dicer* KO cells the decrease in 5mdC is very modest compared to the sustained DNA demethylation of wild type cells of the same background (C57B6) (Fig. 6.2a); thus indicating that miRNAs may be important regulators of 2i-induced DNA demethylation.

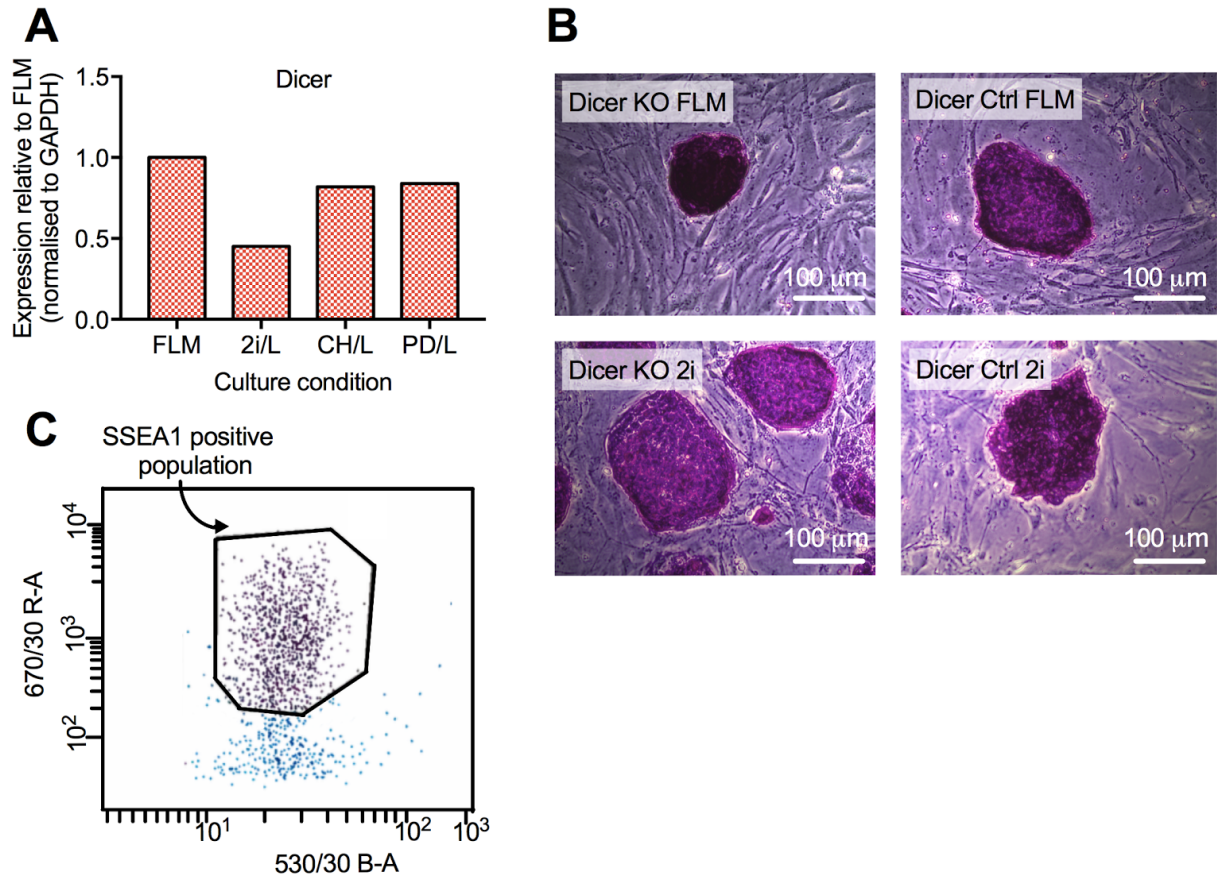


Figure 6.1 Dicer is differentially expressed in FLM and 2i culture conditions

(a) RT-PCR analysis of *Dicer* expression showing that *Dicer* is downregulated in 2i compared to FLM culture conditions. This downregulation depends on dual inhibition by CHIR99021 and PD0325901. Cells were cultured by Dr. H. Leitch whilst gene expression was analysed by Dr. K. McEwen.

(b) *Dicer* KO cells cultured in both FLM and 2i culture conditions are positive for alkaline phosphatase staining. Crucially, *Dicer* KO mESCs can only be grown on highly dense feeder cells.

(c) *Dicer* KO (or Ctrl) mESCs were therefore separated by FACS from the feeder cells based on the expression of the mESC-specific SSEA1 surface marker.

Quantification of epigenetic regulators by RT-qPCR shows that *Dicer* KO mESCs are largely comparable to their genetically matched wild type counterparts; *Dnmt1* expression is unchanged whilst both *Dnmt3b* and *Dnmt3l* are downregulated in 2i (Fig. 6.3b). Crucially however, neither *Dnmt3a* isoform 1 nor isoform 2 are downregulated in 2i (Fig. 6.3a). Therefore, failure to initiate DNA demethylation in *Dicer* KO mESCs may be due to the continued expression of *Dnmt3a*.

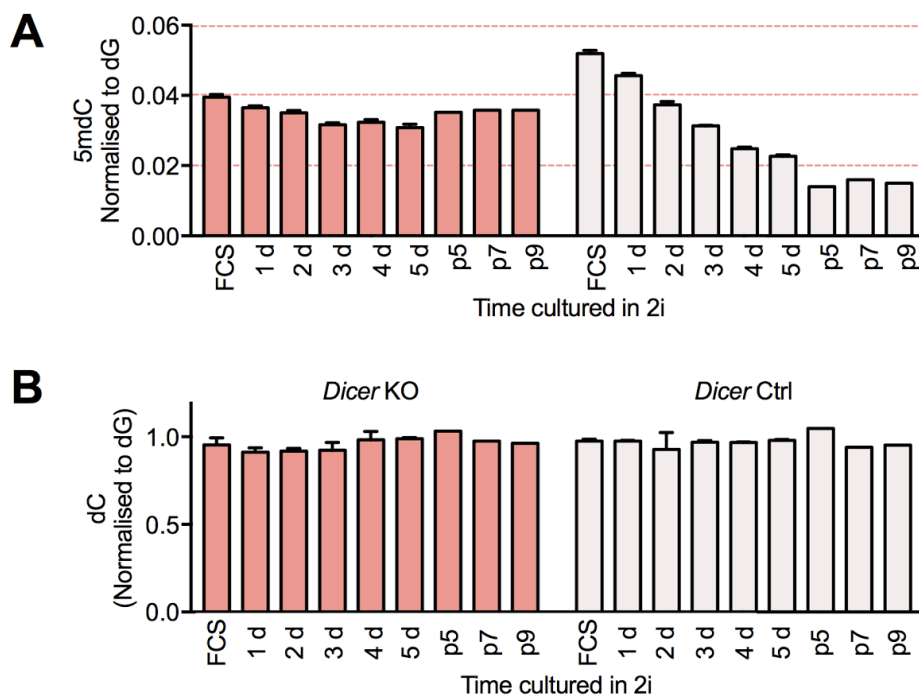


Figure 6.2 LC-MS quantification of 5mdC in Dicer KO and Ctrl cells

(a) Dicer KO and Ctrl cells maintained in FCS + MEF + Lif (FLM) culture conditions were transferred into 2i. Cells were harvested every 24 hours for the first 5 days, and at passage (p) 5, p7 and p9. 5mdC was quantified by LC-MS and the results normalised to the abundance of dG. The dC/dG ratio is also shown as a measure of the reliability of this analysis.

To explore further how miRNAs regulate the transition between FLM and 2i culture-induced gene expression programmes, small RNAs were extracted from wild type mESCs by T. Huang and sequenced using a small RNA-Seq approach. The reads were aligned and normalised to the total reads by Dr. P. Sarkies. miRNAs significantly upregulated or downregulated were defined as having an adjusted p-value of less than 0.01. P-values, expressed as $-\log_{10}(\text{adjusted p-value})$ were plotted against the $\log_2(\text{fold change})$ to aid visualisation of miRNAs with a potentially important role in reprogramming to ground state pluripotency (Fig. 6.4a).

Since the mESC-specific *Dnmt3a* isoform 2 is not downregulated in *Dicer* KO mESCs, we asked whether miRNAs may mediate its post-transcriptional regulation upon transition to 2i. Post-transcriptional regulation of *Dnmt3a* by miRNAs may be direct or indirect (Fig. 6.5). miRNAs predicted (and experimentally validated) to

target the *de novo* Dnmts were identified using the microRNA.org database. Notably, only miRNAs targeting *Dnmt3a* were significantly changed (either up- or down regulated) in 2i (Fig. 6.4a-d).

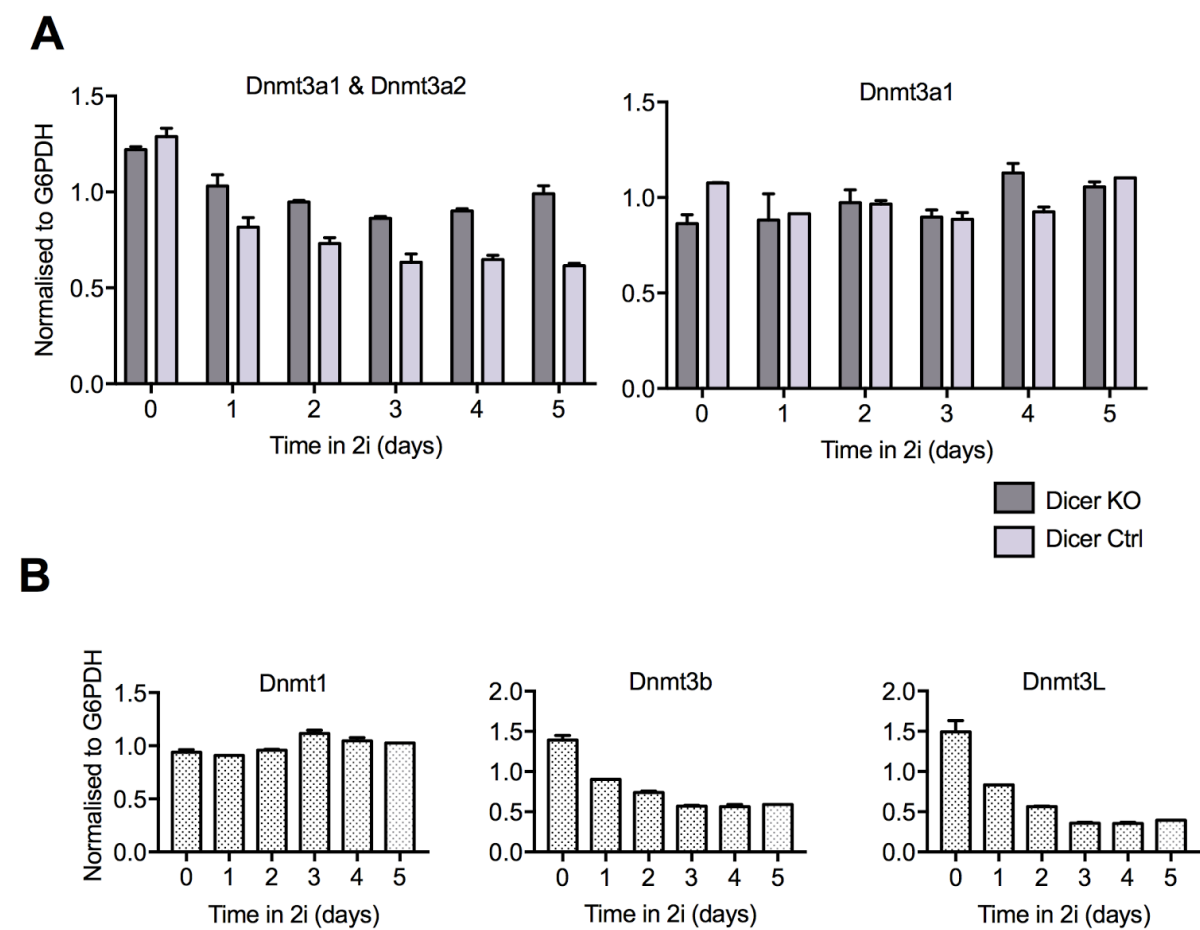


Figure 6.3 Dicer is necessary for Dnmt3a isoform 2 downregulation in 2i
 (a) RT-qPCR analysis of Dnmt3a isoform 1 and isoform 2 expression in Dicer KO cells after transition to 2i.
 (b) All other epigenetic regulators show comparable expression levels between Dicer KO and Ctrl cells.

To determine whether any of these significantly upregulated miRNAs could potentially be involved in the post-transcriptional regulation of *Dnmt3a* in 2i, it will be important to first show that the expression of the miRNA changes rapidly following the transition to 2i conditions. Any miRNA directly targeting *Dnmt3a* must therefore be upregulated in 2i. miRNAs were therefore extracted from wild type mESCs (genetically matched to the *Dicer KO* mESCs) every 12 hours for 72 hours after transition to 2i. The candidate miRNAs expression must now be quantified by RT-qPCR assay. Since miRNAs are only 21 nucleotides long, qPCR analysis relies on first

polyadenylating the small RNAs and then amplifying the miRNA pool. For normalisation, miRNAs known to be expressed uniformly between FCS-growth conditions and 2i have been selected.

6.3.2. miRNA regulation of Dnmt3a isoform 2 transcriptional activators and repressors

As previously shown, *Dnmt3a* isoform 2 (*Dnmt3a2*) is downregulated in wild type mESCs grown in 2i conditions whilst expression of the ubiquitous *Dnmt3a* isoform 1 (*Dnmt3a1*) remains unchanged. These two isoforms are transcribed from separate promoters. Since, miRNAs typically bind to the 3'UTR of mRNAs, any miRNA directly targeting *Dnmt3a2* will inherently also target *Dnmt3a1*. Therefore, rather than targeting *Dnmt3a* directly, miRNAs may differentially regulate *Dnmt3a1* and *Dnmt3a2* by instead targeting *Dnmt3a2* transcriptional regulators. For instance, *Dnmt3a2* expression may be silenced by the upregulation of miRNAs that target *Dnmt3a2* transcriptional activators (Fig. 6.5b) or by the downregulation of miRNAs that target *Dnmt3a2* repressors.

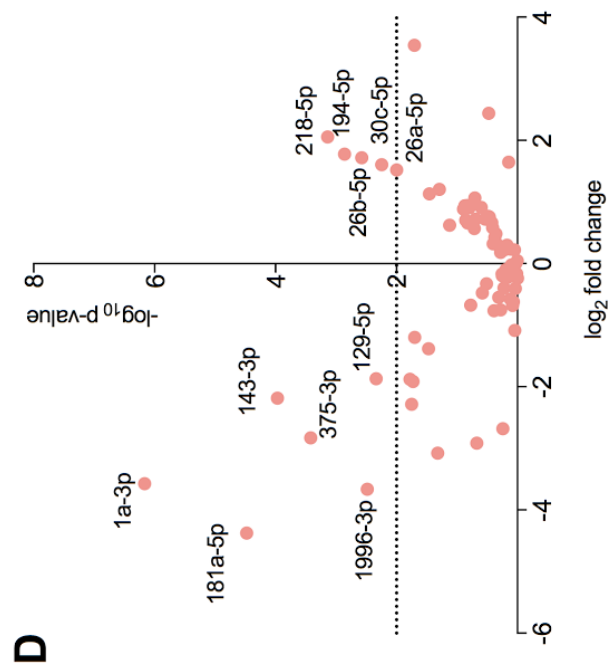
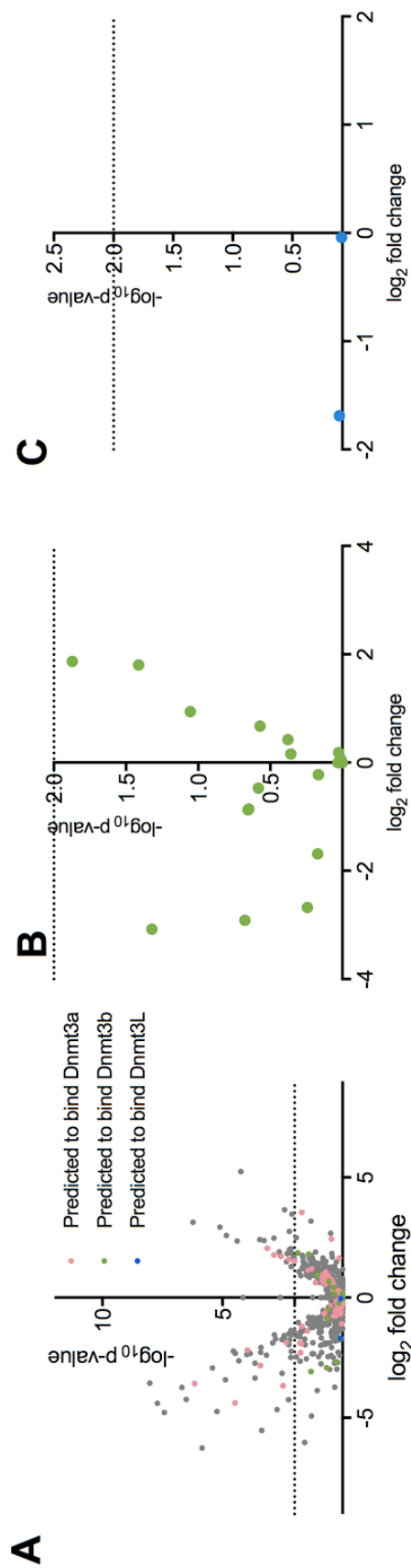


Figure 6.4 miRNAs significantly up- or downregulated in 2i
(a) Volcano plot illustrating the fold change expression and significance of all sequenced miRNAs between FLM and 2i culture conditions. miRNAs predicted to target Dnmt3b (b), Dnmt3l (c) and Dnmt3a (d) were identified using the miRNA.org database. An adjusted p-value of 0.01 was chosen to determine significance

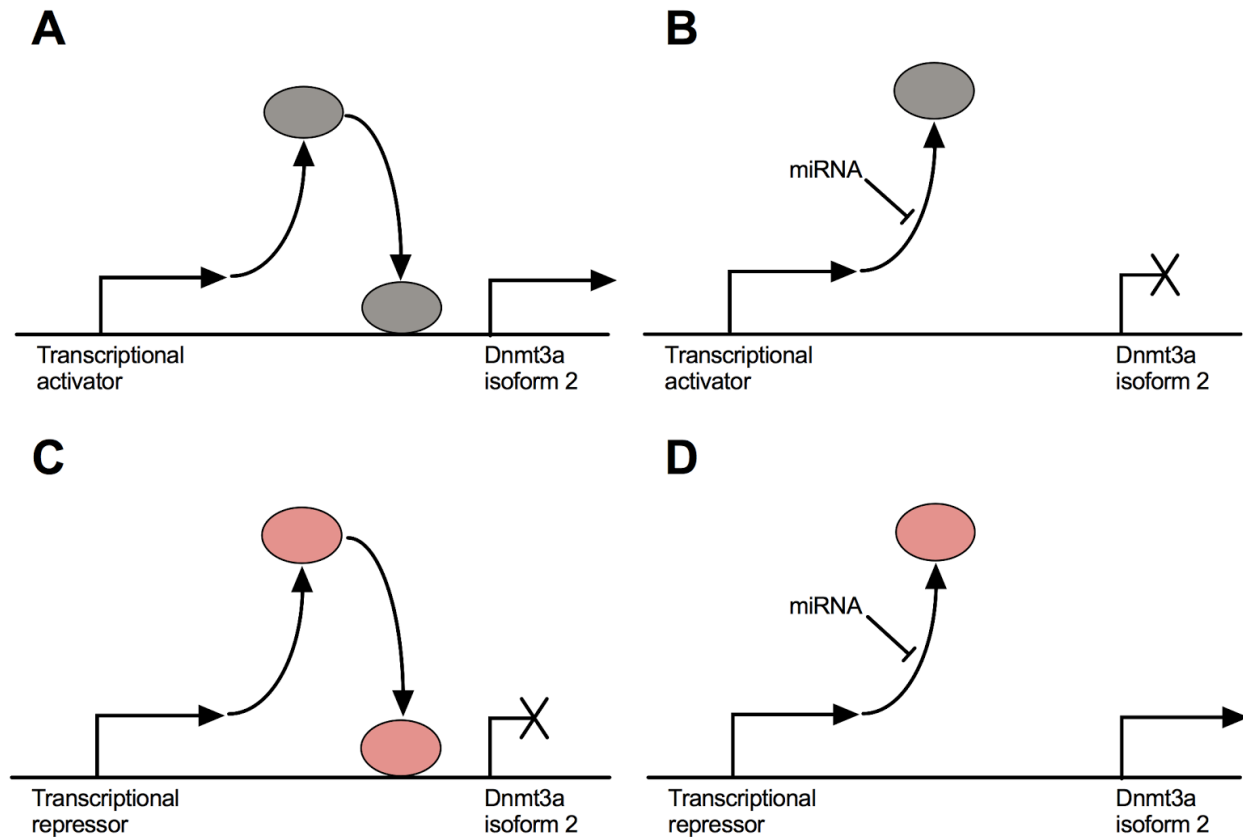


Figure 6.5 Alternative means of miRNA-mediated regulation

- (a) Transcriptional regulators activate Dnmt3a2 transcription.
- (b) miRNAs targeting transcriptional activators indirectly repress Dnmt3a2 expression.
- (c) Transcriptional regulators silence Dnmt3a2 transcription.
- (d) miRNAs targeting transcriptional repressors indirectly activate Dnmt3a2 expression.

Little is known regarding *Dnmt3a2* transcriptional regulation. I therefore initially set out to locate the *Dnmt3a2* promoter. It is known that *Dnmt3a* exons 1 to 6 are only found in *Dnmt3a1*. Exon 7, however is specific to *Dnmt3a2* and the *Dnmt3a2* promoter resides within an intron upstream from exon 7 (Fig. 6.6a). Using the PROSCAN webtool (www-bimas.cit.nih.gov), I identified a 211 bp sequence with a significant probability of being a promoter (promoter score = 86.28) approximately 4 kb upstream from the *Dnmt3a2* specific exon 7 and only 100 bp from the start of the *Dnmt3a2* 5'UTR (Fig. 6.6b). This score reflects the sequence similarity to putative eukaryotic Pol II promoters. Compared to the surrounding sequence, the candidate promoter is highly enriched for E2F and Sp1 binding sites, and to a lesser extent, GCF binding sites.

miRNAs targeting *Dnmt3a2* transcriptional regulators were identified using the miRNA.org miRSVR (miRNA support vector regression) scoring method. Typically, miRNA target sites are predicted using the tripartite miRANDA algorithm based on complementarity between the miRNA and the mRNA 3'UTR, assessing the energetics of the physical interaction between the miRNA and mRNA and scoring the conservation of predicted miRNA:mRNA pairs within closely related species. The miRSVR model uses supervised learning to score miRANDA predicted sites based on mRNA expression changes following miRNA transfection, therefore broadening target prediction beyond the classical notion of complementarity to the seed region. A miRSVR score cutoff of -1.2 was used to assess the likelihood of target downregulation. All but 2 miRNAs with a miRSVR cutoff target E2F5, a transcriptional repressor (Fig. 6.6b).

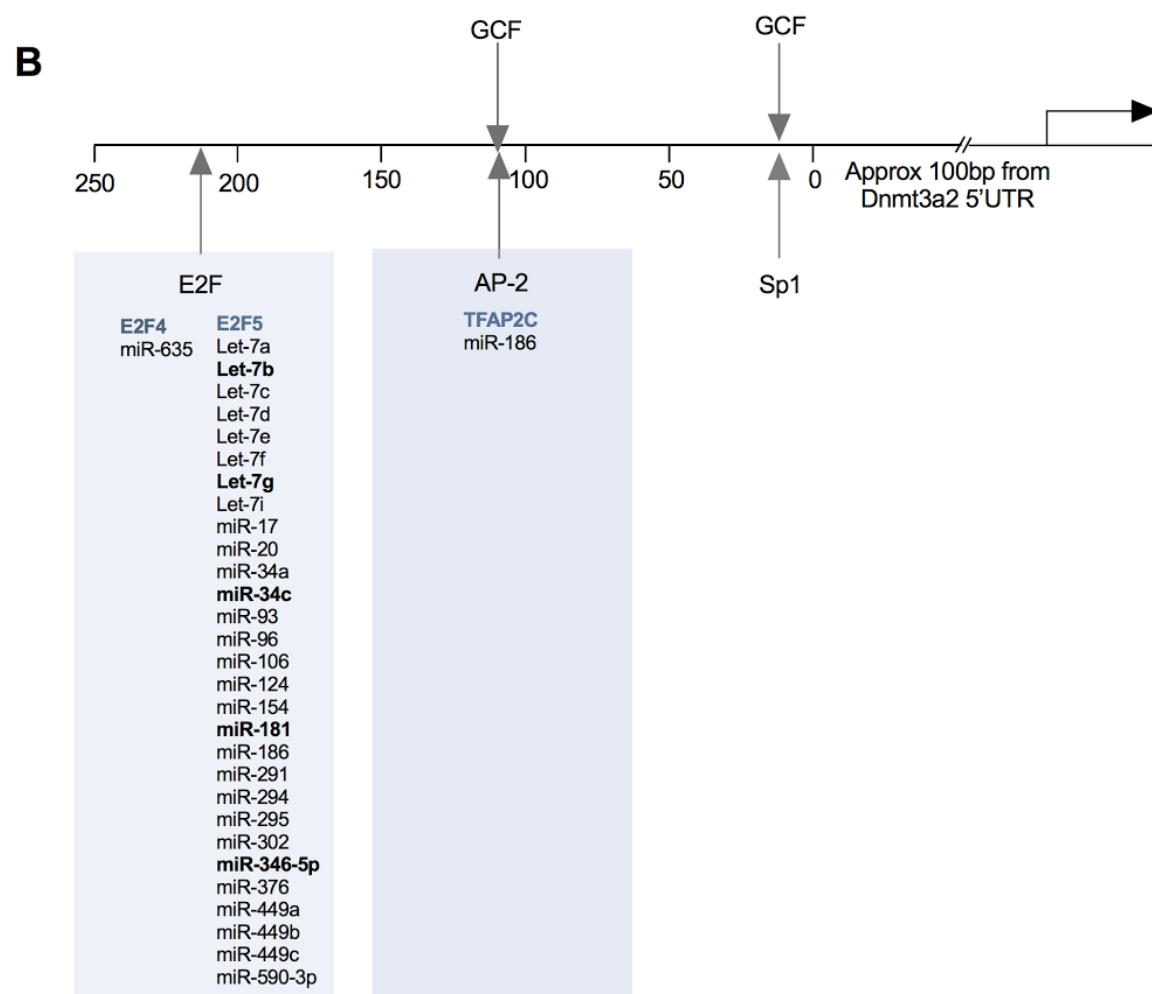
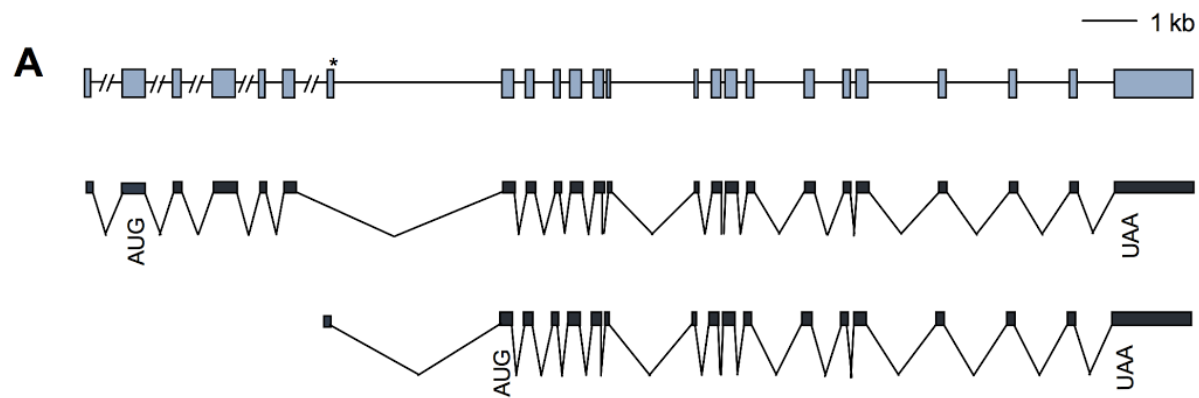


Figure 6.6 miRNA regulation of Dnmt3a isoform 2 expression

(a) Structure of mouse Dnmt3a gene and mRNA. The Dnmt3a2 specific exon is indicated by an astrisk. Transcription of Dnmt3a2 is initiated from a downstream intronic promoter.

(b) The predicted putative Dnmt3a2 promoter based on its sequence similarity to the eukaryotic pol II promoter, located 100bp upstream from the Dnmt3a2 5'UTR. The predicted binding sites of regulatory factors is indicated, along with candidate miRNAs expected to cause their post-transcriptional downregulation. Only miRNAs with a SVR score below -1.2 are shown.

Indirect miRNA-mediated regulation of *Dnmt3a2* expression in conventional FCS or 2i conditions would require the miRNAs to be differentially expressed in each condition. Using the small RNA sequencing results described previously, I selected miRNAs which had an adjusted p-value of less than 0.01 and less than 0.05. These values are visualised in the volcano plot (Fig. 6.7a). Of these miRNAs, the expression of 4 are significantly changed between the culture conditions; Let-7g-5p, miR-34c-5p, miR-181-5p and miR-346-5p. Furthermore, Let-7b-5p is significantly downregulated with an adjusted p-value of less than 0.05.

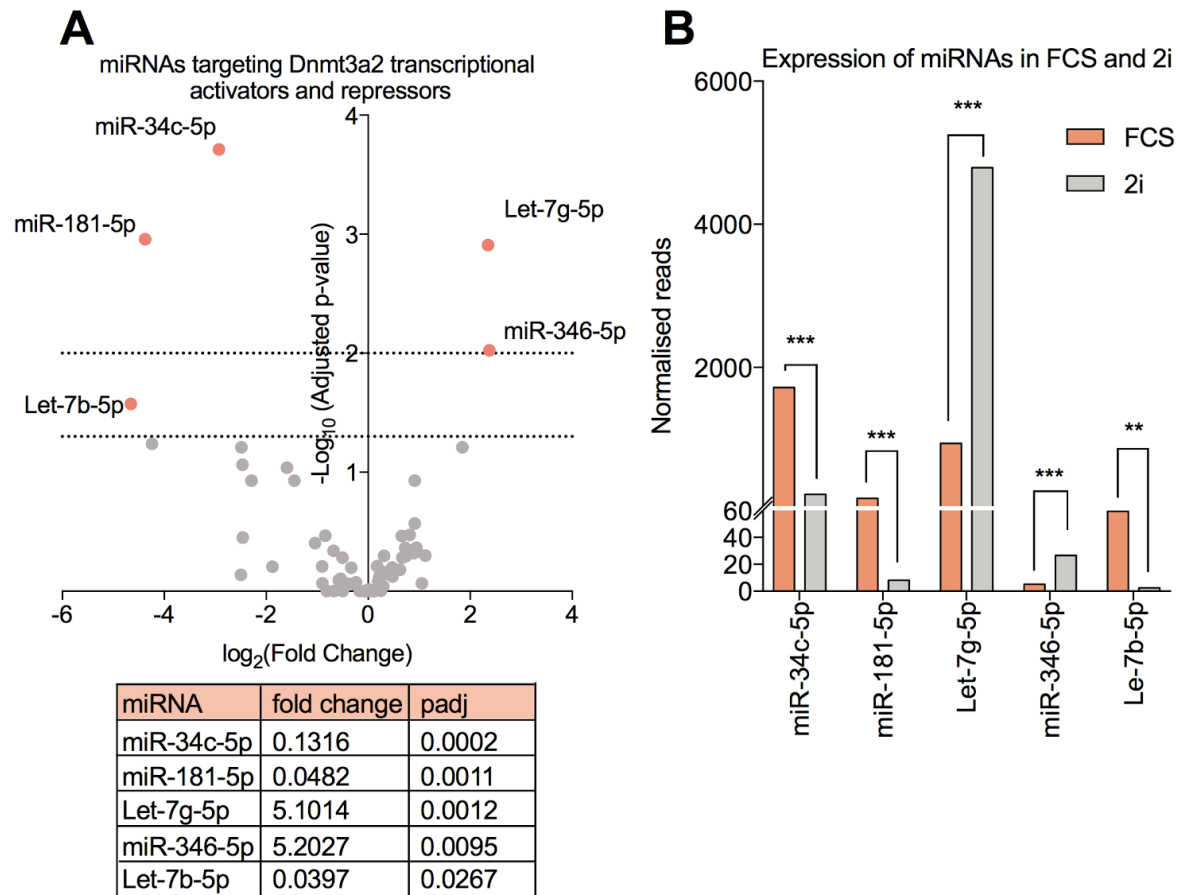


Figure 6.7 Expression of miRNAs targeting Dnmt3a isoform 2 transcriptional regulators

(a) Volcano plot of differentiated miRNAs which are predicted to target Dnmt3a2 transcriptional regulators. miRNAs with an adjusted p-value below 0.05 and 0.01 are highlighted in red.

(b) Average expression of miRNAs in FCS and 2i, normalised to the total reads. * denotes significance: *** P-value < 0.001, ** P-value < 0.01, * P-value < 0.05.

To determine whether these 5 miRNAs are able to indirectly regulate *Dnmt3a2* expression, it must be confirmed that there are also significant changes in expression of the transcriptional regulators between conventional FCS and 2i culture conditions. We subsequently created a RNA-Seq dataset using wild type mESCs maintained in either FLM or 2i culture condition. The normalised reads of AP-2 and E2F transcription factors are shown in Figure 6.8. Importantly, there is no significant difference in expression between E2F transcription factors in mESCs cultured in FLM or 2i conditions. Only AP-2 α is significantly upregulated in 2i. AP-2 α can act as either a transcriptional activator or repressor, and therefore it is possible that AP-2 α specifically may represses

Dnmt3a2 in 2i. However, since miRNAs predicted to target AP-2 α are not significantly changed upon transition to 2i (Fig. 6.6b), this observation cannot explain the downregulation of *Dnmt3a2*.

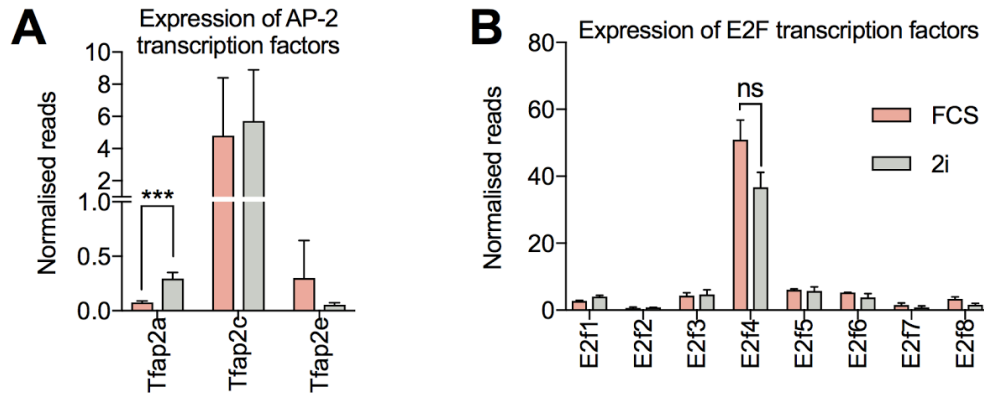


Figure 6.8 Expression of *Dnmt3a* isoform 2 expression in FCS and 2i

(a) Expression of AP-2 transcription in mESCs cultured in FCS and 2i conditions, measured by RNA-Seq

(b) Expression of E2F transcription in mESCs cultured in FCS and 2i conditions, measured by RNA-Seq

*** denotes $p(\text{adj}) < 0.001$

6.3.3. RNA-Seq analysis of *Dicer* KO cells

In addition to regulating the expression of *Dnmt3a*, it is likely that miRNAs are responsible for more widespread changes in gene expression programmes. To assess how miRNAs differentially regulate the states of pluripotency found in FCS and 2i culture conditions, *Dicer* KO mESCs were adapted to 2i + feeder culture conditions and transcriptome analysis was performed using an RNA-Seq approach (Fig. 6.9a). Since mESCs were collected after 5, 7 and 9 passages in 2i, these samples should be considered as biological replicates. As a comparison, two biological replicates were collected for *Dicer* WT and *Dicer* KO cells cultured in FLM. To avoid skewing the analysis due to the presence of partially differentiated cells (especially in the *Dicer* WT sample) or contaminating feeder cells, mESCs were isolated based on the expression of a cell surface protein found specifically on murine pluripotent cells (SSEA1). Therefore, it is important to bear in mind that we are likely to be collecting only a subpopulation of *Dicer* KO and *Dicer* WT mESCs. RNA was extracted from each sample, and cDNA libraries were prepared by Dr. R. Amouroux. Sequenced reads were aligned by Dr. P. Hill.

Morphologically, *Dicer* KO mESCs cultured in FLM and 2i + feeder conditions are very similar. Notably, cells appear to be flatter than their wild type counterparts. However, alkaline phosphatase staining suggests that they retain pluripotency (Fig. 6.1b). Perhaps the most obvious phenotype of *Dicer* KO mESCs is their rate of proliferation, which is substantially slower than wild-type mESCs. Additionally, clonogenicity appears to be reduced; passaging is associated with greater cell death in *Dicer* KO mESCs. Reducing the extent of cell dissociation appears to increase the viability *Dicer* KO mESCs.

6.3.4. Initial exploratory analysis

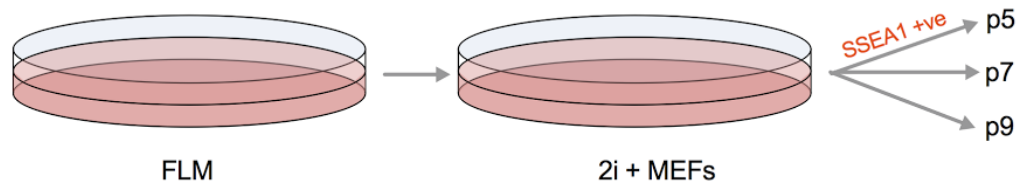
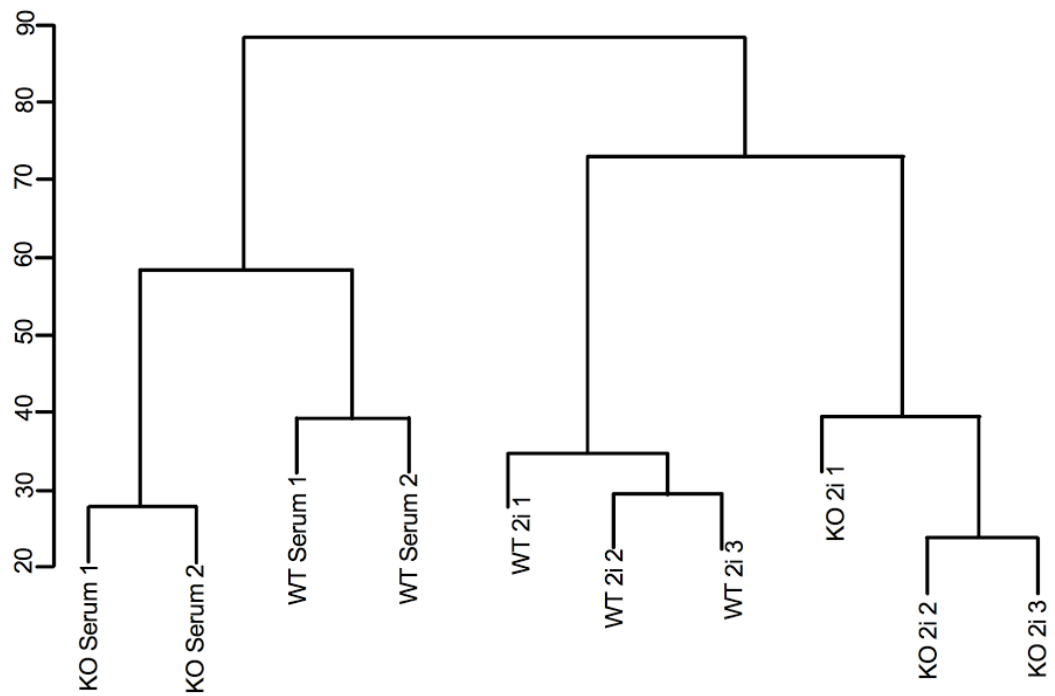
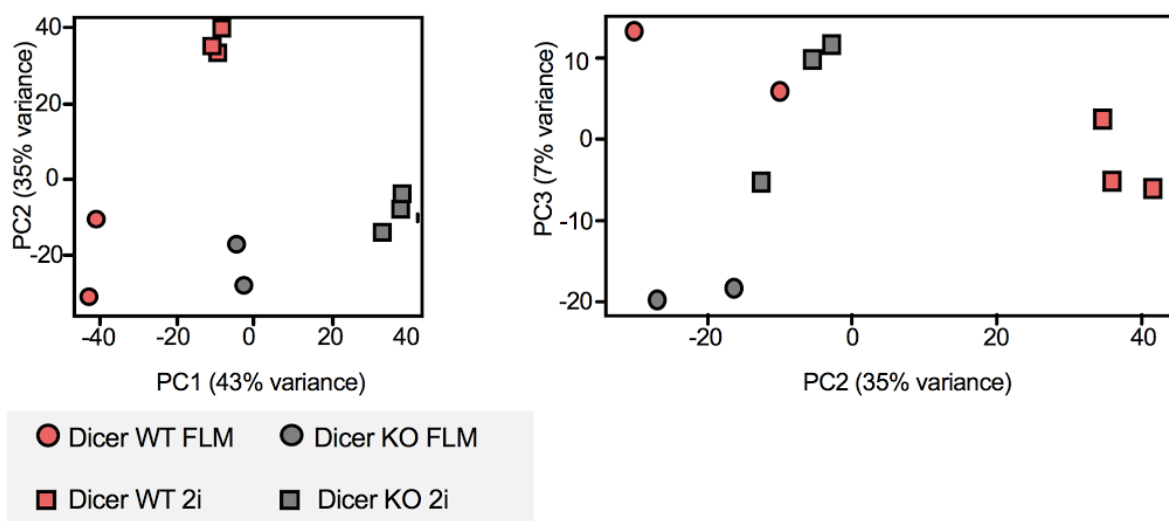
Initially, we looked at the relationship between the sequenced datasets by hierarchical clustering and Principal Component Analysis (PCA) analysis. Hierarchical clustering allows us to build a binary tree based on the similarity of the gene expression profiles between samples. In this instance, the vertical axis represents the distance (or dissimilarity) between clusters and the horizontal axis represents each biological sample (Fig. 6.9b). Therefore, samples with a high degree of dissimilarity will be linked at the top of the dendrogram whilst samples which are more similar will be linked at the bottom. Based on unsupervised hierarchical clustering, biological replicates are highly similar and variation is caused firstly by the culture conditions and secondly by the presence (or absence) of *Dicer*.

Unlike hierarchical clustering, which assumes that each clade (cluster) is homogeneous, Principal Component Analysis (PCA) aims to reduce the dimensionality of a highly complex dataset whilst also retaining most of the datasets variation. PCA identifies the dimensions (principal components) associated with the highest variation between samples, allowing each sample to be represented by several rather than several thousand variables. The PCA plot (Fig 6.9c) suggests that the relationship between cell line and culture condition is more complex than shown by hierarchical clustering. Datasets are separated along PC1 based on their culture condition *and* the presence of DICER. Therefore, even in FLM culture conditions, *Dicer* KO cells are different to *Dicer* WT cells. PC2 illustrates that the cell lines respond differently to the addition of 2i; only *Dicer* WT cells separate along PC2 based on culture condition. Therefore, the initial exploratory analysis aligns with our previous observation,

that *Dicer* KO cells do not undergo 2i-induced global DNA demethylation and transcriptional reprogramming to the ground state. reprogramming to the ground state.

Figure 6.9 Exploratory analysis of Dicer KO RNA-Seq data

(a) Schematic of the experimental system. Dicer KO and Dicer WT cells adapted to 2i + MEF culture conditions were separated from feeder cells based on the expression of SSEA1. RNA extracted from SSEA1 positive cells was used to prepare cDNA libraries for RNA sequencing. (b) Hierarchical clustering of samples. (c) PCA analysis showing the variation between samples.

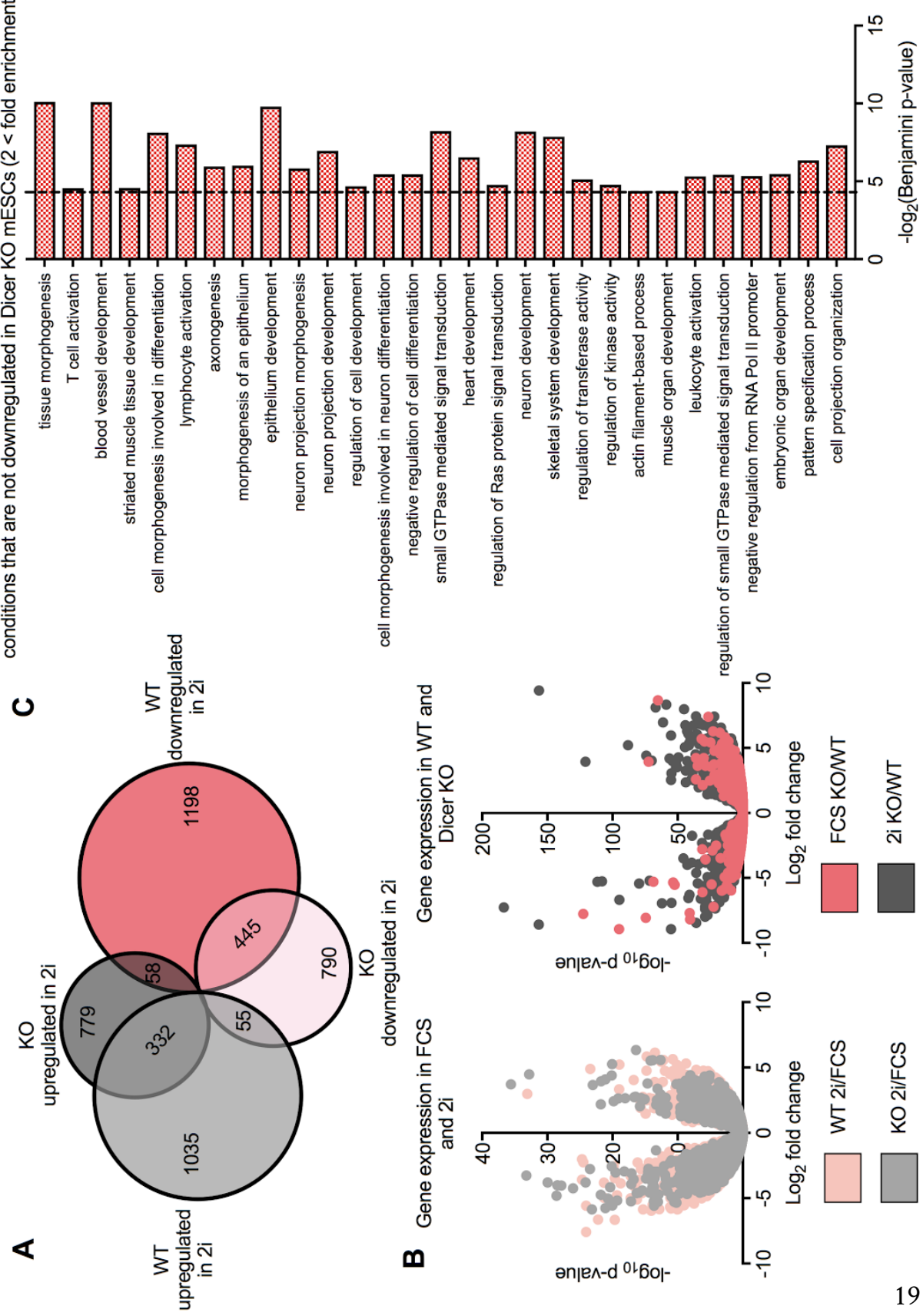
A**B****C**

6.3.5. Genes associated with developmental processes are downregulated in 2i WT and *Dicer* KO cells

Upon transition to 2i, 1035 and 1198 genes are up- or downregulated respectively in WT mESCs. In *Dicer* KO mESCs however, these values are substantially lower; 779 and 790 respectively (Fig. 6.10a). Therefore, in alignment with the PCA analysis, the transcriptional changes induced upon transition to 2i, appear to be far greater for WT mESCs than *Dicer* KO mESCs. The fold change in gene expression caused by the addition of 2i is also greater for WT cells compared to *Dicer* KO cells (Fig. 6.10b). Additionally, when comparing WT and *Dicer* KO cells maintained in the same culture condition, the difference in gene expression is greater in 2i than FLM (Fig. 6.10b).

Gene ontology (GO) analysis shows that the majority of biological processes downregulated in both WT and *Dicer* KO cells cultured in 2i compared to FLM conditions are associated with tissue patterning and development; including cell fate commitment, pattern specification process and skeletal, striated muscle and urogenital development. This is consistent with the known effect of 2i culture conditions. However, GO analysis of genes specifically downregulated in WT cells upon adaptation to 2i conditions (i.e. not downregulated in *Dicer* KO mESCs), indicates that there are many other developmental genes which are not downregulated in *Dicer* KO mESCs after prolonged culture in 2i conditions (Fig. 6.10c-d). GO terms are listed in order of decreasing fold enrichment. Notably, terms associated with GTPase activity, cell motility and the actin cytoskeleton are also specifically downregulated in WT cells cultured in 2i compared to FLM.

Biological processes significantly downregulated in WT mESCs in 2i vs FLM culture conditions that are not downregulated in Dicer KO mESCs (2 < fold enrichment > 2.5)



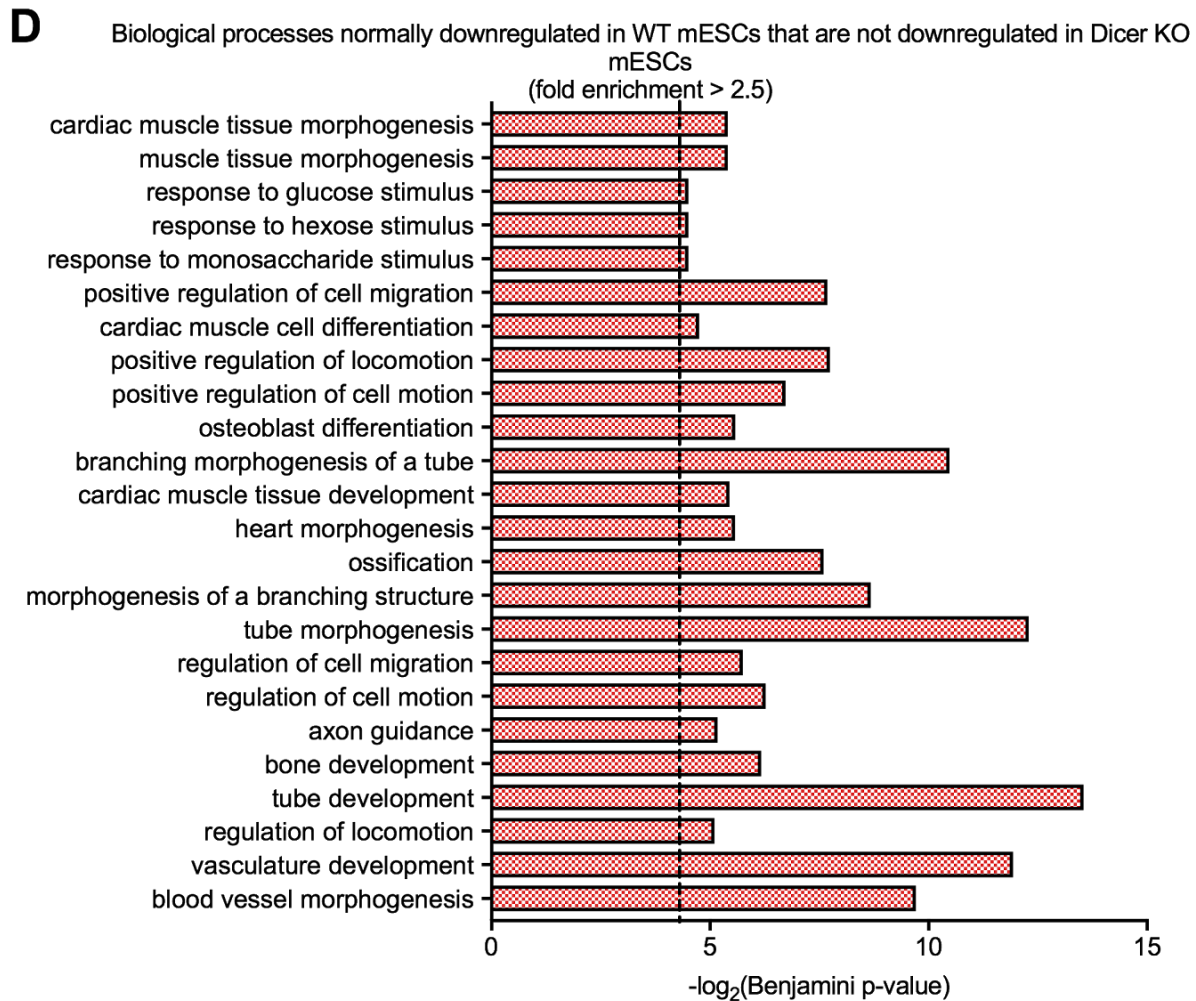


Figure 6.10 Differential gene expression in *Dicer* KO cells cultured in 2i or FLM growth conditions

(a) Venn diagram illustrating the number of genes significantly up- or down-regulated in *Dicer* KO and WT cells upon the transition from FLM to 2i + MEF culture conditions. Significance was defined as having a $p(\text{adj})$ -value < 0.05 and a fold change of between 1.5 and 2.5.

(b) Volcano plots showing the significance and fold change of gene expression within a single cell line when transferred from FLM to 2i + MEF conditions, and between *Dicer* KO and WT cells cultured in either culture condition.

(c) Gene ontology analysis showing biological processes normally downregulated in WT mESCs transferred from FLM to 2i + MEF conditions, that are not downregulated in *Dicer* KO mESCs. Processes are listed in order of increasing fold change. All processes shown are significant. A Benjamini p -value of less than 0.05 was specified.

(d) Gene ontology analysis showing biological processes significantly downregulated in *Dicer* WT cells transferred from FLM to 2i + MEF condition with a fold change of greater than 2.5. Processes are listed in order of increasing fold change.

6.3.6. *Dicer* KO cells show reduced expression of naive pluripotency markers

Previously *Dicer* KO mESCs have been described as hyper-naive and resistant to differentiation. On the other hand, the dysregulation of DNA methylation is known to contribute to stabilisation of the pluripotency circuitry and suppression of differentiation. Therefore, the relationship between miRNAs, DNA methylation and ground state pluripotency is ultimately unknown. Computational modeling has shown that rather than relying on a vast interactome, pluripotency is essentially controlled by 12 components and 16 separate interactions; including the core pluripotency factors and a collection of naive transcription factors (Dunn et al., 2014). OCT4 and SOX2 are indispensable for pluripotency. Stem cell maintenance relies on the finely-tuned expression of each factor; with either under- or overexpression favouring differentiation. The expression of these factor is not altered by the addition of 2i. Furthermore, expression is comparable in WT and *Dicer* KO mESCs (Fig. 6.11b). NANOGs role in pluripotency is more puzzling; deletion does not invariably cause the loss of pluripotency, however *Nanog* expression is necessary to acquire a self-sustaining pluripotent ground state. To this end, in WT cells, *Nanog* expression is significantly upregulated following the addition of 2i. *Nanog* is however highly expressed in *Dicer* KO cells regardless of the culture condition. Speculatively, high *Nanog* expression may contribute to the dysregulation of differentiation observed in *Dicer* KO cells. Ultimately, miRNA depletion does not destabilise the core pluripotency circuit. This is not unexpected since we know that pluripotent *Dicer* KO cells can be maintained *in vitro*.

Naive transcription factors complete the minimal pluripotency network. These proteins are characteristically upregulated in 2i culture conditions (Fig. 6.11b). The expression of most of these remaining factors does not differ significantly between *Dicer* KO and WT mESCs maintained in the same condition; only *Stat3* and *Sall4* expression differ significantly. *Stat3* is more highly expressed in *Dicer* KO cells compared to WT mESCs. The JAK/STAT signalling pathway is known to inhibit the differentiation of mESCs into mesoderm and endoderm (Bourillot et al., 2009). Nonetheless, in *Dicer* KO mESCs, apparent elevation of JAK/STAT signalling coincides with the elevated expression of embryonic endodermal markers (caldesmon 1, *Cald1*), and mesodermal markers (the intermediate filament vimentin, *Vim*), suggesting that the increased activation of the JAK/STAT pathway

and the inhibition of differentiation-inducing pathways may not be coupled. *Sall4* is expressed at significantly lower levels in *Dicer* KO mESCs compared to *Dicer* WT mESCs. Without quantifying the reduced expression of SALL4 at the protein level, it is difficult to assess how transcriptional downregulation may affect the pluripotency circuitry. On the one hand, whilst the remaining pluripotency network may partially compensate for reduced *Sall4* expression, the characterisation as a component of the minimal pluripotency network suggests that reduced *Sall4* activity should be associated with a discernable phenotype. Alternatively, the expression of *Sall4* above a particular threshold may be sufficient to avoid any deleterious phenotype. Finally, *Dicer* KO mESCs may tolerate reduced expression of *Sall4* whilst WT mESCs would not. If true, this would suggest a subtle difference in the wiring of the pluripotency network between WT and *Dicer* KO mESCs.

SALL4 is believed to safeguard self-renewal by buffering mESCs from differentiation inducing signals. This is partially dependent on the recruitment of the NuRD chromatin remodelling complex and histone deacetylase activity to target promoters. For example, NuRD-mediated silencing of *Cdx2* expression represses aberrant trophoblast gene expression (Yuri et al., 2009). Other downstream targets include *Pten*, apoptosis inducing factors (*Tnf*, *Tp53*, *Card9*, *Card11*, *Cysc*, and *Lta*), survival factors (*Bmi-1*, *Bcl2*, *Xiap*, *Dad1*, *Tegt*), the Sonic Hedgehog (SHH) pathway and cell surface molecules including *Epcam* and epithelial-mesenchymal transition (EMT) related factors. The survival factors *Bmi-1*, *Bcl2*, and *Xiap* are significantly upregulated in *Dicer* KO cells whilst the pro-apoptotic *Card9*, *Card11*, and *Cysc* are downregulated in *Dicer* KO cells compared to WT cells; suggesting that reduced *Sall4* expression is potentially felt at the level of cell survival (Fig. 6.11d). Given that *Dicer* KO cells simultaneously show heightened expression of developmental genes and pluripotency factors (*Nanog*), the upregulation of these pathways may facilitate the survival of cells which are barraged by multiple, conflicting signalling pathways.

Figure 6.11 miRNAs and the control of the pluripotency transcriptional network

- (a) Schematic showing the minimal pluripotency network. Inputs (LIF or inhibitors) are shown in grey, factors that antagonise the network are shown in red and pluripotency factors are in black.
- (b) Expression of the minimal pluripotency network, quantified by RNA-Seq, and normalised to the total reads.
- (c) Expression of lineage markers, quantified by RNA-Seq and normalised to the total reads.
- (d) Expression of factors, targeted by SALL4 and involved in the regulation of apoptosis and cell survival.

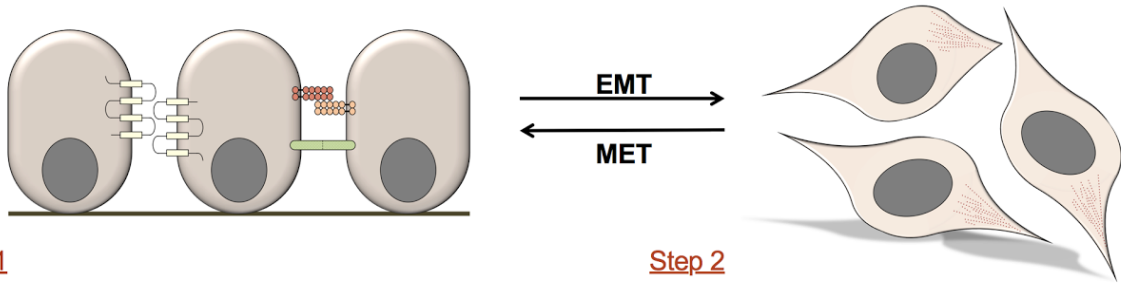
*** denotes p(adj)-value < 0.001, ** p(adj)-value < 0.01, * p(adj)-value < 0.05.

6.3.7. *Dicer* KO cells show characteristics of the epithelial to mesenchymal transition

In addition to reduced GTPase activity, pluripotent cells are also characterised by reduced actomyosin fibres and thus reduced cell viscosity and elasticity; two characteristics necessary for cell motility. These traits are shared by epithelial cells and are typically lost during the epithelial-mesenchymal transition (EMT); a phenomenon characterised by the restructuring of the actin cytoskeleton, the loss of epithelial markers (specifically cell surface adhesion molecules and polarity complexes), the loss of cell-cell adhesive contacts and the upregulation of the mesenchymal gene expression programme. Since cell-cell adhesion and apical-basal polarity facilitate cell differentiation, EMT is also associated with reduced ability to differentiate and increased cell motility, and invasive properties with resistance to apoptosis (reviewed in Chen et al., 2017).

Embryonic stem cells and their epiblast counterparts are epithelioid in nature. However during gastrulation, many of these cells differentiate into the mesoderm. Therefore, differentiation of mESCs is likened to the EMT (Chen et al., 2017). *Dicer* KO cells also display many EMT traits. Significantly, this includes the loss of E-cadherin (*Cdh1*) (Fig. 6.13a), which is highly expressed in WT mESCs maintained in both FLM and 2i culture conditions. E-cadherin expression is regulated by transcription factors including TWIST, SNAIL, SLUG, ZEB2 and TCF3 (Cano et al., 2000). Both *Snail* and *Zeb* are more highly expressed in *Dicer* KO cells compared to WT mESCs; with *Zeb2* being barely detectable in ground state WT mESCs maintained in 2i. A reciprocal upregulation of mesenchymal genes is seen, including N-cadherin (*Cdh2*), the intermediate filament protein vimentin (*Vim*), and the mesenchymal cell surface protein NCAM (Fig. 6.13b). Additionally, transcripts for matrix metalloproteinases including gelatinase (*Mmp2* and *Mmp9*) are more abundant in *Dicer* KO cells (Fig. 6.13b).

A



Step 1

Disassembly of epithelial cell-cell contacts including tight junctions, adherens junctions, gap junctions and desmosomes

Loss of cell polarity through disruption of CRUMBS, PAR and SCRIB complexes

Activation of mesenchymal gene expression

Step 2

Reorganisation of the actin cytoskeleton

Expression of matrix metalloproteinases

B

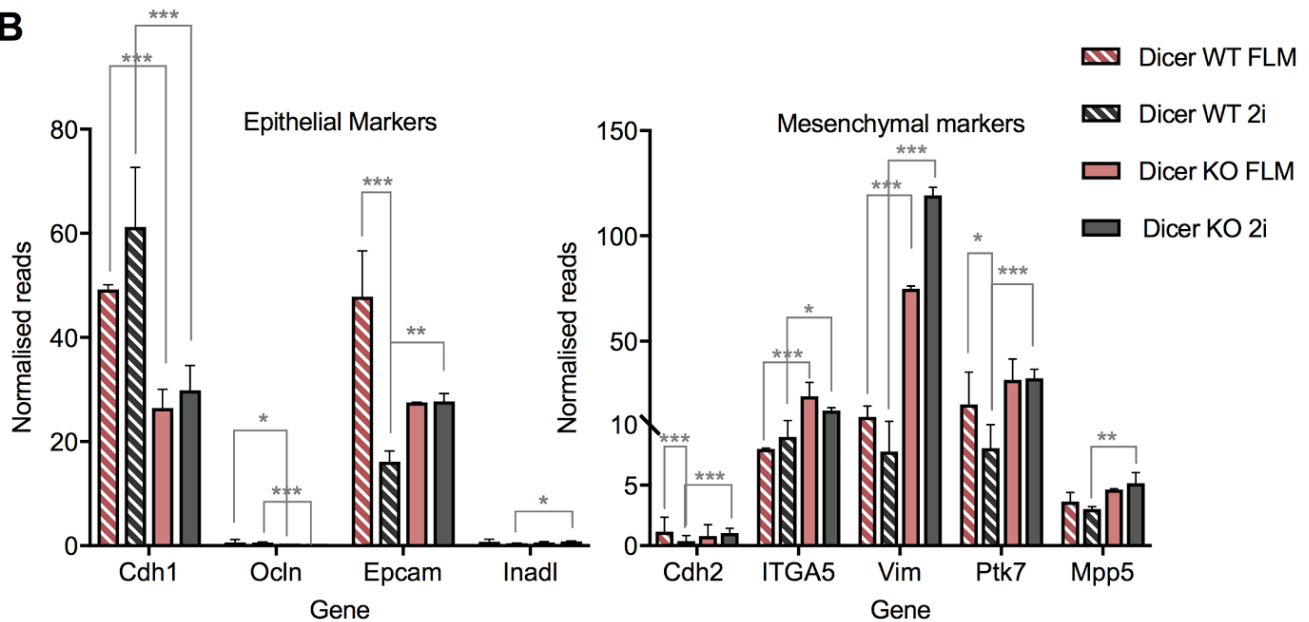


Figure 6.12 miRNAs and the epithelial-mesenchymal transition

(a) Schematic illustrating the characteristic cellular changes associated with the epithelial-mesenchymal transition.

(b) Expression of epithelial and mesenchymal markers, quantified by RNA-Seq and normalised to the total reads.

*** denotes $p(\text{adj})$ value < 0.001 , ** $p(\text{adj}) < 0.01$ and * $p(\text{adj}) < 0.05$.

6.3.8. Stat3/Jak signalling and 2i inhibition alone cannot support the survival of *Dicer* KO cells

We have observed that regardless of culture condition, maintenance of *Dicer* KO mESCs relies on the use of a highly confluent feeder layer (personal communication with Dr T. Nesterova, Oxford University). Increased cell viability may result from paracrine signalling (MEFs are a rich source of NODAL, FGF4, BMPs and EGF) or mechanical support. Since mESCs grown in 2i should not strictly require LIF, this raises questions over the role of E-cadherin and STAT signalling in *Dicer* KO cells.

E-cadherin is essential for LIF-dependent mESC self renewal and prevents differentiation into EpiLCs; genetic ablation has shown that *Cdh1* is required for STAT3 activation upon ligand binding to LIFR (Larue et al., 1994). Culture of E-cadherin null cells in the presence of BMPs partially restores naivety by inducing the expression of N-cadherin (*Cdh2*), and therefore providing additional adhesive contacts (Hawkins et al., 2012). Since E-cadherin forms a ternary complex with LIFR and gp130, it has been suggested that its primary function is to stabilise the LIFR-gp130 complex. Together, these observations may explain why feeder-free conditions cannot support the maintenance of *Dicer* KO cells. Alternatively, since Rho activation and Rac inhibition underlie dissociation the induced apoptosis of hESCs (Ohgushi et al., 2010), *Dicer* KO cells (and hESCs) may require increased contact with adjacent cells to counteract their own elasticity. To test whether actin polymerisation has a detrimental effect on *Dicer* KO phenotype, cells could be cultured in the presence of the ROCK inhibitor Y-27632 in FCS and FLM conditions.

6.3.9. miRNA and epigenetic regulation of ground state pluripotency

In agreement with the RT-qPCR analysis presented earlier, whilst all *de novo* DNA methyltransferases are transcriptionally downregulated in WT mESCs upon transition to 2i, in *Dicer* KO mESCs *Dnmt3a* transcriptional expression is not altered by dual GSK3 β and ERK inhibition. Furthermore, in 2i culture conditions the expression of *Dnmt3l* is significantly lower in *Dicer* KO cells compared to WT cells. Since reduced

expression of *Dnmt3l* is likely to impair any residual *de novo* DNA methyltransferase activity, it is perhaps surprising that *Dicer* KO mESCs do not become hypomethylated in 2i culture conditions.

One of the prevailing mechanisms for 2i-triggered DNA demethylation is based on PRDM14-mediated repression of the *de novo* Dnmts. It is therefore interesting that *Dnmt3b* and *Dnmt3l* are nevertheless downregulated in *Dicer* KO mESCs without concurrent *Prdm14* upregulation. Based on the results described in the this chapter, it seems plausible that *Dnmt3a* is regulated post-transcriptionally by miRNAs. Since we have previously seen that 2i primarily causes the downregulation of the ESC-specific *Dnmt3a* isoform 2, it will be necessary to re-align the RNA-Seq reads to take into account the two differentially expressed *Dnmt3a* isoforms. Finally, in agreement with our RT-qPCR, *Dnmt1* expression in *Dicer* WT and *Dicer* KO cells does not change significantly upon transition to 2i.

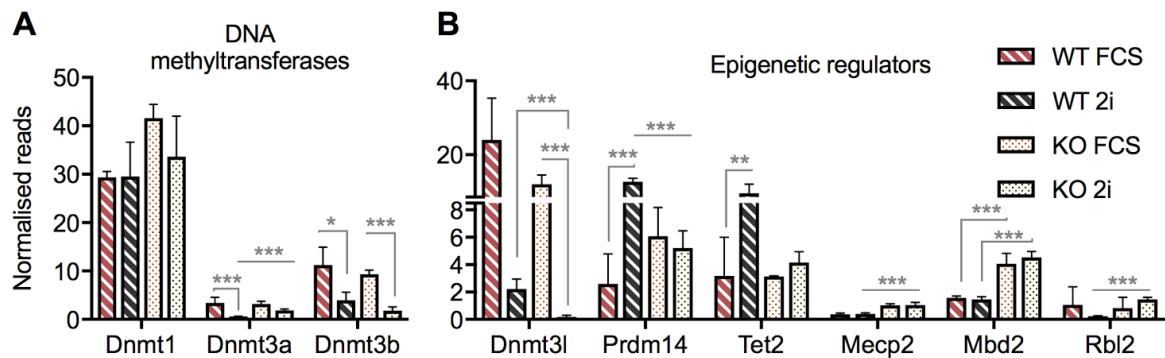


Figure 6.13 Expression of epigenetic writers and readers mediated by post-transcriptional regulation

(a) Expression of DNA methyltransferases in *Dicer* KO and WT cells in FLM and 2i culture conditions, quantified by RNA seq and expressed as the number of normalised reads. Expression of *Dnmt1* does not significantly change. Upon transition to 2i, *Dnmt3a* is downregulated in *Dicer* WT mESCs only. *Dnmt3b* is significantly downregulated in both *Dicer* KO and *Dicer* WT cells in 2i.

(b) Expression of epigenetic regulators.

*** denotes $p(\text{adj})$ value < 0.001 , ** $p(\text{adj})$ value < 0.01 , * $p(\text{adj})$ < 0.05 .

6.4 Discussion

We know that miRNAs are important regulators of pluripotency, but little is known about whether they differentially regulate mESCs maintained in FCS and 2i culture conditions, and whether they play a role in 2i-induced DNA demethylation. We have shown that *Dicer* KO mESCs in FCS-based medium are hypomethylated compared to mESCs of the same background. However, they also appear to be unable to undergo DNA demethylation when transferred to 2i medium.

Since only *Dnmt3b* and *Dnmt3l* are downregulated in *Dicer* KO mESCs upon transition to 2i, increased levels of 5mdC in *Dicer* KO mESCs is likely to be caused by the continued transcriptional expression of *Dnmt3a* isoform 2. Additionally, although not quantified directly, *Dicer* KO mESCs proliferated considerably slower than their wildtype counterparts, and could easily be passaged for 5 days without splitting. Dysfunctional cell cycle regulation is known to be a phenotype of *Dicer* loss, and it is possible that this contributes to the relative DNA hypermethylation in *Dicer* KO mESCs.

Previously, *Dicer* KO mESCs have been described as being both hypo- or hyper-naïve. It is true that when cultured in FLM conditions, *Dicer* KO mESCs express higher levels of *Nanog* and are hypomethylated compared to WT mESCs of the same genetic background. Additionally, based on our Principal Component Analysis, *Dicer* KO mESCs cultured in FLM are most similar to WT cells cultured in 2i + MEFs. However, upon transition to 2i, they do not lose 5mdC and based on our PCA, the difference between *Dicer* KO mESCs cultured in FLM and 2i + MEFs is smaller than that seen between *Dicer* WT cells cultured in FLM or 2i + MEF conditions. In fact, in both FLM and 2i + MEF culture conditions, *Dicer* KO mESCs express significantly higher levels of lineage markers and Hox genes. This upregulation does not appear to be specific to a particular lineage. Finally, the classical signs of EMT and the upregulation of *Xist* are instead characteristic of more differentiated cells. Therefore, it might be that *Dicer* KO mESCs represent an epigenetic state that lies somewhere between the ground state maintained in 2i, and mESCs maintained in FLM conditions. *Dicer* KO mESCs, however are

effectively immobilised in this one state; being unable to reprogram to the ground state whilst also being unable to initiate a coordinated programme of differentiation. Therefore, semantically, it seems incorrect to classify them as naive in any respect. Importantly, all cells used for this analysis were SSEA1 positive. SSEA1 is a marker of pluripotent stem cells and is involved in the adhesion and migration of cells in the preimplantation embryo. Therefore, it is possible that we are looking at only a subpopulation of mESCs. These populations may be different if miRNAs regulate SSEA1 expression in mESCs.

Signaling pathways are highly dose dependent. Therefore, conceptually, it is easy to envisage that miRNAs could define a particular signalling landscape. Cells are bombarded by many different, often conflicting extracellular signals and must be able to determine which signals to respond to. miRNAs are likely to play a key role in defining a cell's response to these extracellular cues. On the one hand, miRNAs may actively repress signalling pathways, defining a threshold below which the cell will not respond and therefore enabling the cell to determine which signals require a response, and which signals are merely weak or transient background noise. Specifically, Wnt, Notch and hedgehog signalling are known to be controlled by miRNA mediated default repression. Conversely, by targeting inhibitors of signalling pathways, miRNAs can amplify signals to create, for example, sharpened morphogenic gradients during embryogenesis. Crucially, miRNAs also link different signalling pathways, facilitating a coordinated cellular response. This is important in generating a robust and reproducible response to a given stimulus and may explain why we detect the transcriptional upregulation of STAT, PI3K, NOTCH and HIPPO signalling pathways.

The number of miRNA families correlates with biological complexity. Thus it seems likely that miRNAs are integral to the interpretation and response to a given differentiation inducing stimuli. In the absence of miRNAs, we see an upregulation of components of many different signalling pathways, including those relating to pluripotency and lineage commitment or differentiation. It is possible, that differentiation relies on a coordinated response to these extracellular stimuli; with pathways leading towards differentiation into a specific lineage upregulated whilst those preventing differentiation, or promoting differentiation into a different lineage

are repressed. Therefore, it is difficult to characterise these cells as being similar to a particular physiological or developmental state. They are clearly a unique, artificial system.

Since loss of miRNAs may affect the fine tuning of multiple signalling pathways and gene expression programmes, it would be interesting to see whether miRNA depletion affects gene expression variability. Typically, high expression variability is thought to reflect technical artifacts. However, biological variance itself can be a useful tool in identifying the mechanisms underlying phenotypic differences. Both genetic and phenotypic variation is significant for a population's adaptability and thus evolutionary fitness. High variance, indicative of a gene with fewer constraints, is associated with increased expression (and perhaps developmental) plasticity. Changes in variance can have opposing effects. On the one hand, reduced variance and therefore less plasticity within a network would reduce a cell's ability to respond to aberrant signalling. Alternatively, reduced variance can increase the predictability of signalling network outcomes. This concept can be visualised using Conrad Waddington's depiction of the canalisation of developmental states, in which high transcriptional variability would result in decanalisation or flattening of the developmental landscape, whilst reduced transcriptional variability would result in deeper, increasingly robust developmental pathways. It could be envisaged that highly variable gene expression profiles caused by an inability to buffer environmental or genetic fluctuations could flatten this landscape to the extent where differentiation is completely dysregulated. The increasing use of single cell technologies could be a useful means of probing the transcriptional variability within a single Dicer KO mESCs.

It has been suggested that DNA demethylation in mESCs transitioned to 2i medium phenocopies epigenetic reprogramming in migratory PGCs. Since PGCs can only be obtained in small numbers, this system may be used to generate hypotheses and guide our *in vivo* studies. For example, the work presented in this chapter may be a useful starting point for the study of how miRNAs regulate DNA demethylation in migratory PGCs. It is currently known that the miR-290/295 and miR-17-92 clusters are highly expressed in PGCs. By comparing the expression of miRNAs in mESCs cultured in FCS and 2i medium, we may identify further miRNAs with a role in PGC development. Ultimately, it is likely that numerous mechanisms are responsible for the transformation

of the epigenetic landscape upon transition into 2i medium. The challenge is likely to be understanding to what extent each pathway contributes to this change. This picture is invariably complicated by culture-defined changes to the cells metabolism.

Chapter 7

The pleiotropic effects of retinol and its derivatives
on pluripotency

7.1 Introduction

7.1.1. Retinoic acid and development

All trans retinoic acid (ATRA) is a potent, active derivative of liposoluble vitamin A (retinol). Through diffusion (morphogen-like) gradients and compartmentalisation caused by retinol metabolism and ATRA catabolism, retinoic acid signalling instructs the commitment of unspecified precursors towards different cell fates, thereby mediating tissue patterning, cell proliferation and apoptosis (Duester, 2008)(Kumar and Duester, 2014). Retinoid homeostasis is critical for both embryonic and adult development; with both excess and deficiencies resulting in a range of congenital abnormalities (Sinning, 1998)(Maden, 2000). The pathways mediating retinoic acid (RA) signalling during development are highly conserved between vertebrate species and centre on the switching of retinoic acid receptors from transcriptional repressors to transcriptional activators.

7.1.2. Retinoid metabolism

The bidirectional flux of vitamin A between the extracellular and intracellular compartments is facilitated by the transmembrane RA-inducible protein Stra6 (fig.7.1) (Kawaguchi et al., 2007)(Amengual, 2014). In the cytoplasm, vitamin A is oxidised to retinaldehyde by alcohol- and retinol dehydrogenases. ADH5 is ubiquitously expressed throughout the embryo and adult whilst expression of ADH1 and 7 is tissue restricted (Ang, 1996). This process is reversible. The second step, the oxidation of retinaldehyde to retinoic acid by retinaldehyde dehydrogenases (Raldh), is irreversible. Raldh1, Radhl2 and Radhl3 display distinct and specific expression patterns; with Raldh2 being initially expressed in the primitive streak and mesodermal cells (Niederreither et al., 1997)(Zhao et al., 1996). Accumulation of ATRA is reduced by oxidative metabolism. Specifically, the cytochrome P450 26 subfamily catalyse the conversion of ATRA into more polar substrates. These can be further oxidised to 4-oxo-RA. This is important in preventing potentially detrimental teratogenic effects of RA

signalling. Administering ATRA *in vivo* and *ex vivo* has been found to lead to the rapid upregulation of Cyp26a1 (Loudig, 2005), suggesting that retinoic acid signalling is constrained by autoregulatory negative feedback loops.

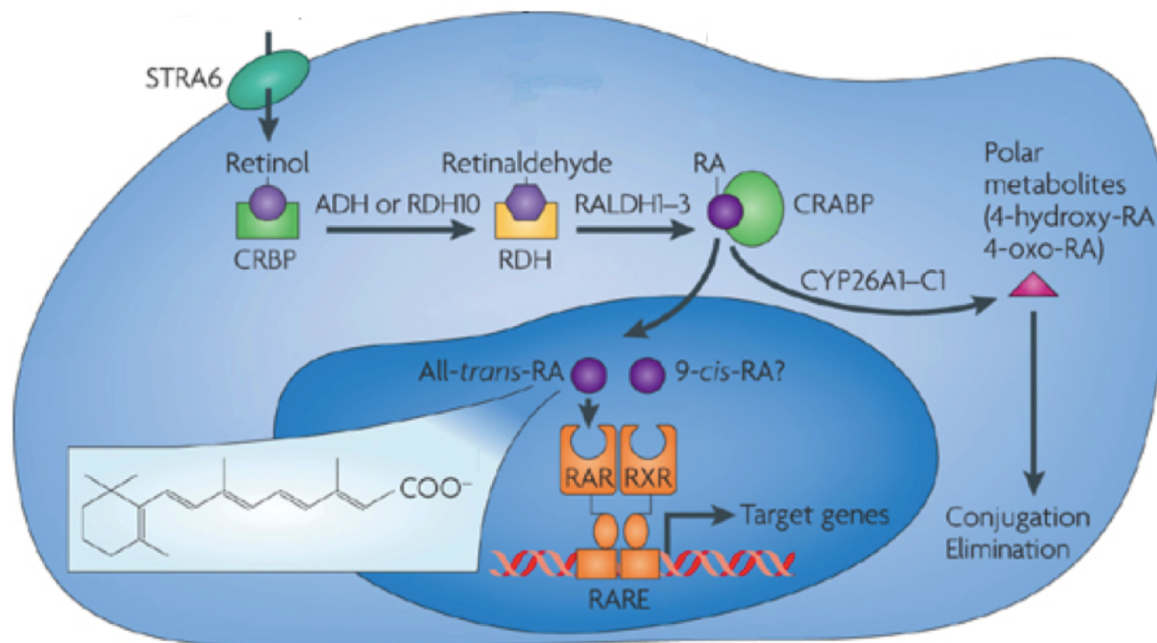


Figure 7.1 Retinoic acid signalling

Retinol is transported into the cell by the receptor stimulated by retinoic acid (STRA6). Retinol is converted to retinoic acid in a two-step reaction; catalysed firstly by retinol dehydrogenase and secondly by retinaldehyde dehydrogenase. All-trans (or 9-cis) retinoic acid (ATRA) is transported to the nucleus by cytoplasmic retinoic acid binding proteins (CRABP) where it binds heterodimeric complexes of the retinoic acid receptors (RARs and RXRs). RAR/RXR complexes bind to retinoic acid receptor elements to activate target gene expression.

Adapted from (Niederreither & Dolie et al., 2008)

7.1.3 Mechanism of retinoic acid signalling

Cytoplasmic ATRA binds the intracellular lipid binding protein (iLBP), cellular retinoic acid binding proteins II (CRABP II), and translocates to the nucleus. Interestingly, the related CRABP I sequesters ATRA within the cytoplasm, where it is targeted for degradation. During embryonic development CRABP I and II are differentially expressed. In the nucleus, ATRA is delivered to its cognate nuclear receptor, retinoic acid binding protein (RAR). RARs are highly conserved throughout vertebrates and form heterodimeric complexes with the more ubiquitously expressed RXRs. Since RXRs bind the ATRA stereoisomer 9-cis-RA (which is not detected endogenously in embryonic or adult tissue) they are suggested to act as scaffolding protein to aid DNA binding

(Chawla et al., 2001). RAR/RXRs constitutively bind retinoic acid response elements (RAREs). In the absence of a ligand RARs/RXRs recruit corepressor complexes (NCoR1 and NCoR2) which deacetylate histones, inducing chromatin compaction and maintaining target gene repression (Balmer et al., 2002)(Chambon et al., 1996)(Lefebvre et al., 2005). Conformational changes associated with ligand binding release the corepressor in favor of binding to the p160 subfamily of steroid receptor coactivators (NCoA1, NCoA2 and NCoA3); leading to the remodelling of chromatin surrounding the RARE and facilitating the assembly of the transcription pre-initiation complex. Numerous genes are activated by RA signalling; among the most important being genes from within the retinoid pathway (*Rarb*, *Crabp1/2* and *Cyp26a1*) and several members of the Hox gene family (Balmer and Blomhoff, 2002).

7.1.4. Retinoic acid, embryogenesis and ESC differentiation

The teratogenic effects of global and local ATRA elevation highlights the importance of controlled retinoic acid signalling during embryogenesis (Ross et al., 2000)(Avantaggiato et al., 1996)(Durst et al., 1989). Due to the ability to modulate cell differentiation programmes in a time and concentration dependent manner (Rohwedel et al., 1999), ATRA is key for the directed differentiation of embryonic stem cells into primitive endoderm-like cells (Sherman and Miller, 1978), neural cells (Maden and Holder, 1992), skeletal and vascular smooth muscle cells, adipocytes, pancreatic β cells (Shi et al., 2005) and cardiomyocytes (Wobus et al., 1997) (Rohwedel et al., 1999). The pleiotropic effects of retinoic acid signalling are partially tissue specific; possibly reflecting modulation by other signalling pathways (FGF, Hedgehog, TGF β). For instance, during hindbrain patterning, FGF prevents the ATRA induced expression of *Cyp26a1*, thereby acting as a feedforward mechanism to enhance the effects of retinoic acid signalling whilst somitogenesis is controlled by antagonising WNT/FGF and ATRA gradients (del Corral et al., 2003).

7.1.5. Retinoids and pluripotency

More recently, the goal of unifying the ATRA regulatory network has been complicated by studies suggesting that retinoic acid signalling may also be active in undifferentiated, pluripotent stem cells. High concentrations (2-3.5 μ M) of vitamin A have been shown to increase hESC proliferation and expression of pluripotency markers (Nanog and Oct4) and stem cell markers (TRA-160 and SSEA4) (Rajala et al., 2011). Similarly, ATRA heightened the proliferative capacity of haematopoietic stem cells whilst simultaneously enhancing the differentiation of downstream progenitors (Khillan et al., 2014). Cancer stem cells (CSCs) react similarly, with vitamin A maintaining the growth from mouse mammary tumours whilst eliminating non-CSCs (Sharma et al., 2013). Finally, forced expression of RARs (Rarg and Lrh-1) or ATRA agonists promoted the reprogramming of somatic cells by the *Oct4* locus (Wang et al., 2011). In mESCs ATRA treatment is followed by Wnt3a upregulation (Martinez-Ceballos et al., 2005). Since, WNT3A-conditioned medium promotes human and mESC self-renewal (Sato et al., 2004) and Wnt/ β -Catenin/CBP signaling maintains mESC pluripotency (Miyabayashi et al., 2007) this may provide a mechanistic explanation for ATRA-induced pluripotency. Finally, transient endogenous ATRA has been detected in the embryo prior to implantation (Rossant et al., 1991; Parrow et al., 1998; Clagett-Dame and DeLuca, 2002). Thus, an appreciation of the role of retinoic acid signalling in pluripotent stem cells could help explain this observation.

7.1.6. Non-canonical retinoid signalling

It is increasingly appreciated that RARs have additional, non-nuclear effects, including the activation of the MAPK pathway (Al Tanoury, 2013). This is thought to involve an extranuclear pool of RARs. In ESCs, for instance RARs present in lipid rafts activate PI3K signalling pathway. Alternatively, vitamin A itself may have retinoic acid-independent functions in mESCs (Chen et al., 2007)(Chen et al., 2008)(Chen et al., 2010)(Zhang et al., 2011)(Sharma et al., 2013). For instance, it has been suggested that vitamin A can potentiate mESC self renewal by elevating Nanog expression independently of the signalling networks classically associated with pluripotency (JAK/STAT, BMP2/4 and WNT/ β -Catenin) (Chen et al., 2007)(Chen et al., 2008). Instead,

vitamin A may act as a growth factor, directly activating PI3K via the IGF-1 receptor, initiating the activation of kinase cascades including phosphorylation of Akt at positions Thr-308 and Ser-473. Ultimately, this is thought to activate mTORC1 and mTORC2 (Chen et al., 2010) and elevate the expression of Nanog and Oct4. The mTOR pathway integrates nutritional and mitogenic cues to coordinate cell growth and the cell cycle; thus potentially linking vitamin A and cell proliferation. For instance, mouse GSCs show increased proliferation and upregulation of genes involved in cell cycle progression including PCNA, c-Myc, cyclin D1, cyclin A and CDK2) when exposed to vitamin A (Wrobel et al., 1996)(Conrad et al., 2008). Interestingly, mESCs do not express many of the enzymes involved in vitamin A uptake or metabolism (including Adh1, Adh2, Raldh2, Stra6 and CRBP), supporting a role for an extracellular growth factor-like function of vitamin A. This may also explain the differential response of pluripotent stem cells and lineage specified cells to vitamin A; vitamin A treatment of mouse ESCs and iPSCs has been shown to create a pure population of undifferentiated cells (Bhatia et al., 2012), possibly due to elimination by terminal differentiation of any cells able to produce ATRA.

7.2 Experimental aims

In addition to dual inhibition of GSK3 β and Erk, two defined supplements are added to basal 2i culture media; N2 and B27. These supplements contain antioxidants and free radical scavengers, insulin, transferrin and albumin. B27 is also commercially available with or without vitamin A derivatives. Vitamin A has three active forms; vitamin A, retinal and retinoic acid. Of these, retinoic acid is the most potent metabolite with key roles in development and the ability to promote mESC differentiation. Therefore, B27 without vitamin A is advertised, and used by many laboratories, for the long term maintenance of naive ground state embryonic stem cells. However, counterintuitively, we noticed that mESCs cultured in 2i medium without vitamin A had higher global 5mdC compared to mESCs cultured in 2i medium with vitamin A. We therefore wanted to investigate the mechanism driving vitamin A induced DNA demethylation.

7.3 Results

7.3.1. Vitamin A enhances DNA demethylation during epigenetic reprogramming to the naive ground state

We observed that adaption of cells to 2i with or without vitamin A had profound effects on DNA methylation and the ability to reprogramme mESCs to the naive ground state. Specifically, DNA demethylation in 2i *without* vitamin A is more gradual and ultimately less pronounced than DNA demethylation in 2i *with* vitamin A. This is surprising given the association between retinoic acid and loss of pluripotency. Certainly, by appearance alone, mESCs maintained in 2i *without* vitamin A have all the characteristics of ground state pluripotent cells; including dome-shaped colonies and clonogenicity.

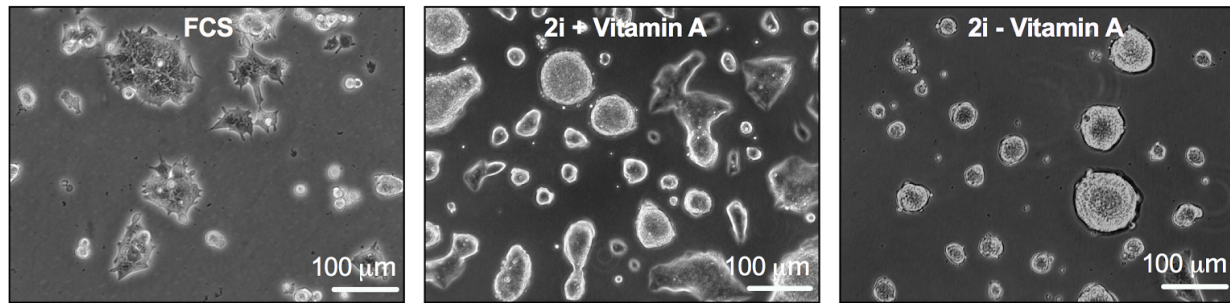
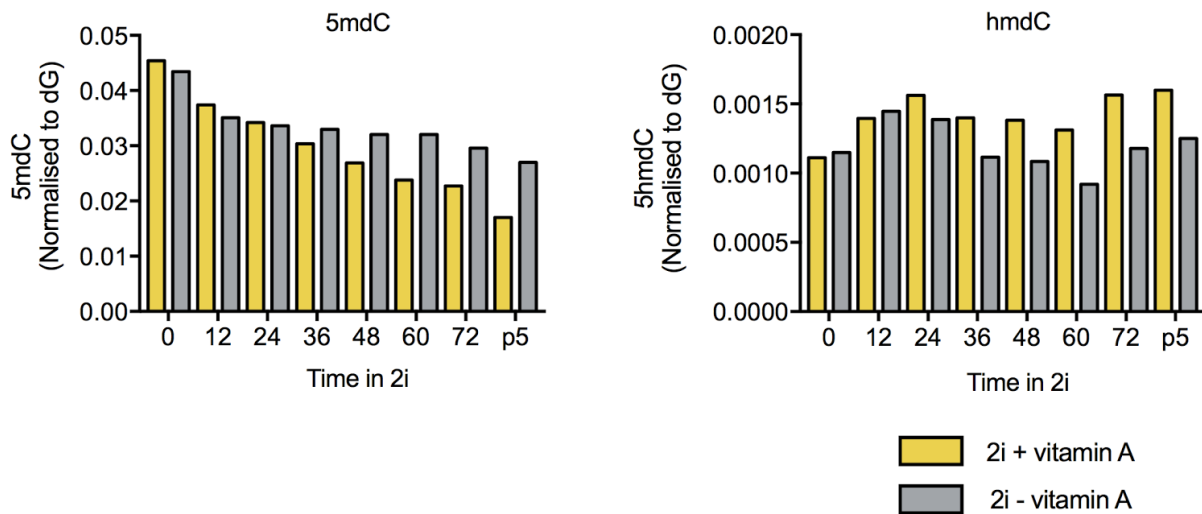
A**B**

Figure 7.2 Effect of vitamin A on cell morphology and DNA modifications

(a) Micrographs showing the typical morphology of mESCs maintained in FCS and 2i (*with* and *without* vitamin A) culture conditions. Notably, cell colonies grown *without* vitamin A show a more uniform, rounded morphology but are substantially smaller than colonies grown in medium supplemented with vitamin A.

(b) LC-MS quantification of 5mdC and 5hmdC abundance during the adaptation of mESCs to 2i. The charts show the mean and SEM of three technical replicates.

7.3.2. Retinol is the active metabolite

The precise composition of the B27 supplement is unknown and in house formulations use a variety of vitamin A derivatives, including vitamin A, retinyl acetate and retinoic acid. To determine which derivative contributes towards extensive DNA demethylation, mESCs were passaged in either 2i/LIF medium *without* vitamin A, supplemented with vitamin A, or retinyl acetate. Of these conditions, only 2i/LIF supplemented with vitamin A

supported the sustained and extensive DNA demethylation necessary to erase 75% global 5mdC (Fig. 7.3), suggesting that vitamin A per se is necessary for extensive DNA demethylation. It is however, important to note that the biological effect of vitamin A or its metabolites is likely to be highly concentration dependent. Therefore, to confirm that vitamin A is the active metabolite, this experiment must be repeated using a range of different vitamin A, retinyl acetate and retinoic acid concentrations.

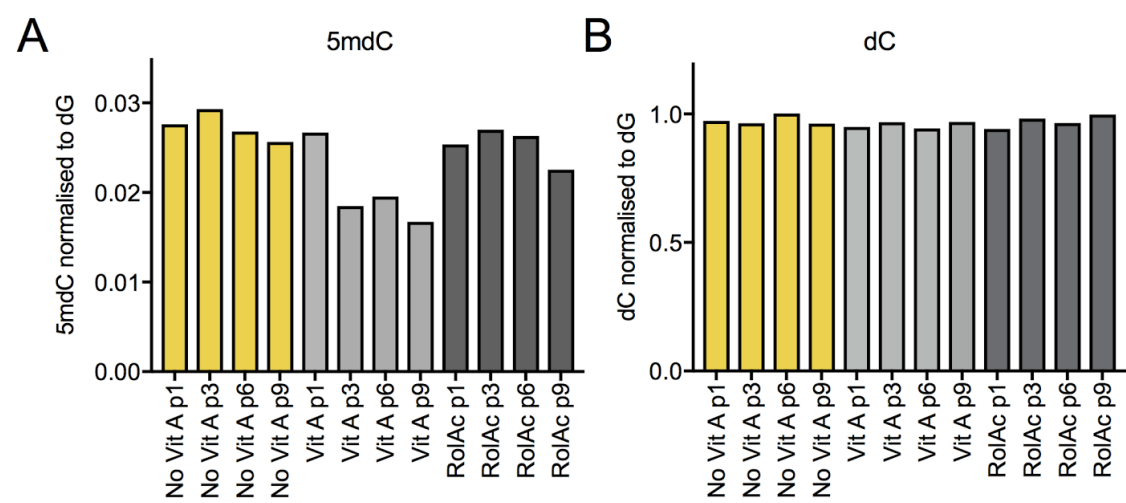


Figure 7.3 LC-MS analysis of the effects of retinol (Vit A) on epigenetic reprogramming

(a) mESCs (12g4) maintained in 2i *without* retinol (Vit A) were transferred to growth medium supplemented with either vitamin A (commercially available), or retinyl acetate (Dr. H. Leitch). Cells were collected after 1, 3, 6 and 9 passages in each condition and the abundance of 5mdC was quantified by LC-MS. The chart shows the mean global 5mdC level (normalised to dG) of three technical replicates. (b) The dC/dG ratio for each sample is also shown as an indicator of technical variability and accuracy.

RT-qPCR analysis of all cells grown in the presence of retinol (vit A) derivatives showed that the presence of vitamin A increases the expression of *Prdm14* while the expression of both maintenance and *de novo* Dnmts is unchanged. Only *Dnmt3l* appears to be downregulated in the absence of vitamin A, and it is unclear whether or not this slight decrease can account for the pronounced hypomethylation in 2i plus vitamin A conditions.

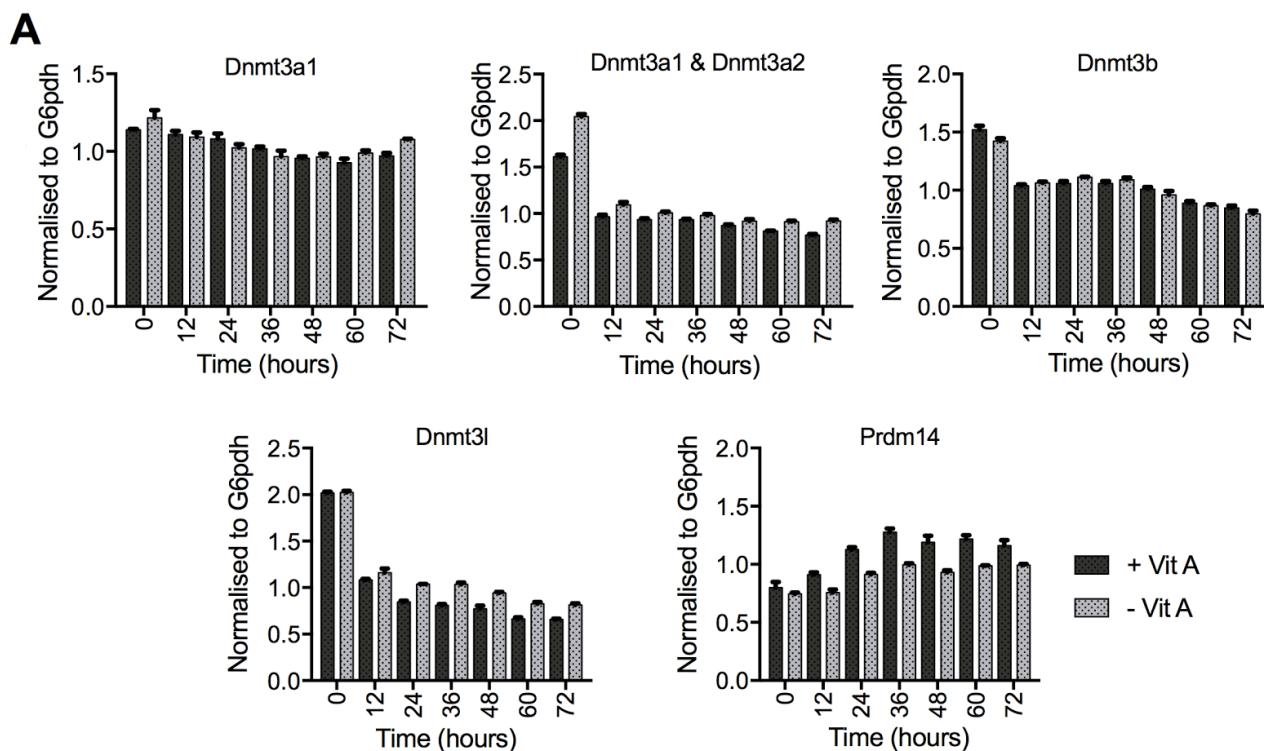


Figure 7.4 Effect of vitamin A on the transcriptional expression of DNA methyltransferases and their regulators.

(a) RT-qPCR analysis of *Dnmt3a* isoform1, *Dnmt3a* isoform 1 and isoform 2, *Dnmt3b*, *Dnmt3l* and *Prdm14* in 12g4 mESCs transitioned from FCS culture conditions to 2i (+ or - vitamin A) conditions. Relative expression was quantified using a standard curve approach. All inputs were normalised to the expression of *G6pdh*. The above results show the mean input and standard deviation between three technical replicates.

Prdm14 links the regulation of the epigenetic ground state and the pluripotency network (Yamaji et al., 2013). Therefore, it is not surprising that pluripotency master regulators (*Oct4*, *Sox2* and *Nanog*) are significantly upregulated in 2i + vitamin A conditions (Fig. 7.5). This is accompanied by an increased expression of naive transcription factors (*Klf4* and *Tbx*) which are exclusively expressed in naive mESCs rather than primed mESCs. Therefore, acquiring ground state pluripotency is crucially dependent on the presence of vitamin A in the B27 supplement. This difference is not widely appreciated and may explain some of the inconsistencies in the data surrounding pluripotency in 2i.

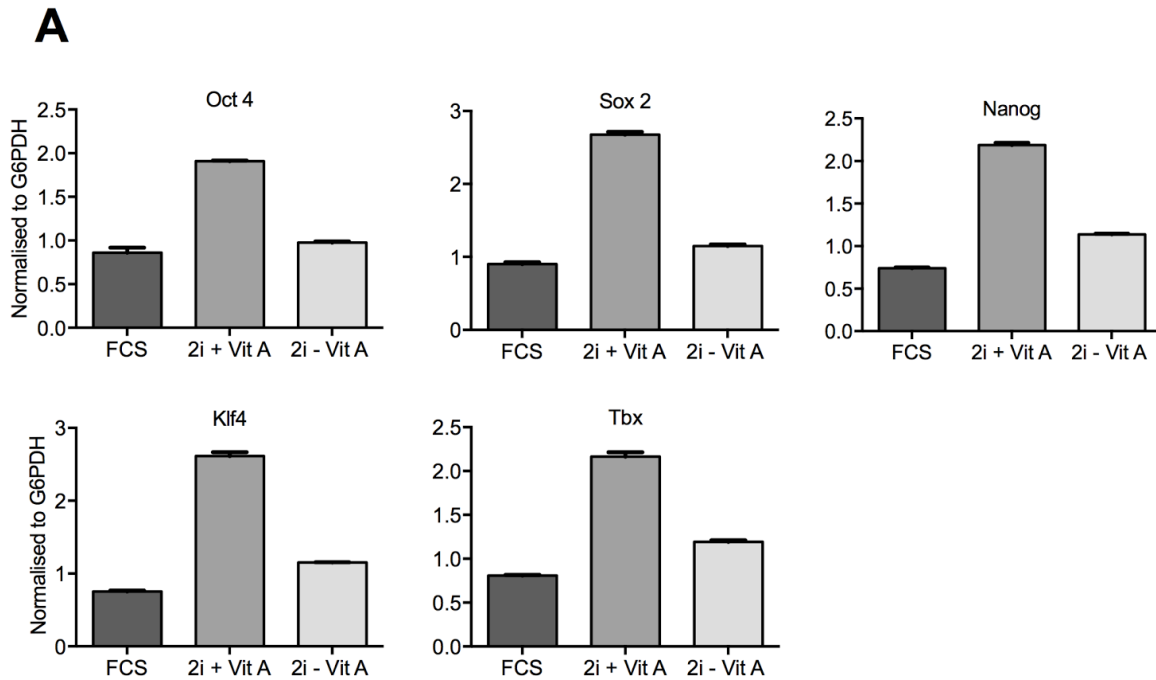


Fig. 7.5 The presence of vitamin A causes an upregulation of the pluripotency transcriptional circuitry
(a) RT-qPCR analysis of pluripotency marker and naive transcription factor expression in mESCs (12g4) maintained in FCS, 2i with vitamin A, or 2i without vitamin A for a minimum of 5 passages. Samples were analysed in triplicate. Relative inputs were quantified using a standard curve. All values are normalised to *G6pdh* expression.

7.3.3. Expression of the ATRA signalling pathway is not upregulated in 2i *with* vitamin A

Using the RNA-Seq dataset described in Chapter 6, we quantified the expression of enzymes involved in retinoic acid signalling in wild type mESCs cultured in FCS and 2i (with vitamin A) conditions. Retinol (vitamin A) is transported into the cell by the transmembrane retinol receptor and cytokine signalling transporter, *Stra6* (Chen et al., 2010). Expression of *Stra6* is known to increase during ATRA-induced differentiation. Notably, *Stra6* expression is significantly downregulated in 2i *with* vitamin A compared to FCS culture conditions; suggesting that vitamin A uptake and retinoic acid signalling is not activated following the supplementation of B27 with vitamin A.

The metabolism of vitamin A to ATRA depends on two successive oxidation steps; Adh 1, 3 and 4 catalysed formation of retinal followed by oxidation to ATRA. Typically, Adh1 and 4 expression is tissue specific whilst expression of Adh3 is ubiquitous. Nonetheless, only Adh1 appears to be expressed in ESCs grown in either conventional FCS or 2i with vitamin A; albeit at very low levels in FCS. Alternatively, oxidation may be catalysed by the microsomal enzymes RDH1, RDH5 and RDH11. Of these, only RDH11 is expressed in mESCs. Therefore, assuming that vitamin A is able to traverse the cell membrane, the subsequent production of retinoic acid is unlikely to be efficient. Retinal dehydrogenases catalyse the second oxidation step. Several isoforms play a role in ATRA production; ADLH1A1, ADLH1A2, ADLH1A3, and ALDH8A1. Of these, only Aldh1a2 is significantly upregulated in 2i; although it is difficult to appreciate the biological effect of this change if vitamin A uptake and retinal formation is severely dysregulated.

CYP26a1, CYP26b1 and CYP26c1 metabolise ATRA to more polar derivatives. The expression of all three genes does not differ between conventional FCS and 2i *with* vitamin A. CYP26 expression is dependent on RAR/RXR heterodimers and RA signalling, therefore suggesting that ATRA signalling is not increased in the presence of vitamin A. Finally, the nuclear retinoid receptors RAR α and RXR α are significantly downregulated in 2i *with* vitamin A. Collectively the expression pattern of these key components of ATRA signalling suggest that it is not upregulated in the presence of vitamin A.

We used western blot to quantify the expression of the nuclear receptors (RARs and RXRs) at the protein level in mESCs cultured in conventional FCS, 2i *plus* Vitamin A and 2i *minus* Vitamin A. Interestingly, at the protein level, there appears to be a slight upregulation of RARs. Although these results were confirmed by immunofluorescence staining and confocal microscopy, somewhat surprisingly, the receptors were not nuclear. Therefore, the upregulation of RARs may have a role beyond their canonical involvement in ATRA signalling. Similarly, expression of the cytoplasmic binding protein CRABP-II, which typically shuttles cytoplasmic ATRA to nuclear RAR/RXR heterodimers, is also upregulated. How the upregulation and cytoplasmic localisation of these receptors can be reconciled with the low expression of enzymes capable of metabolising vitamin A is not known.

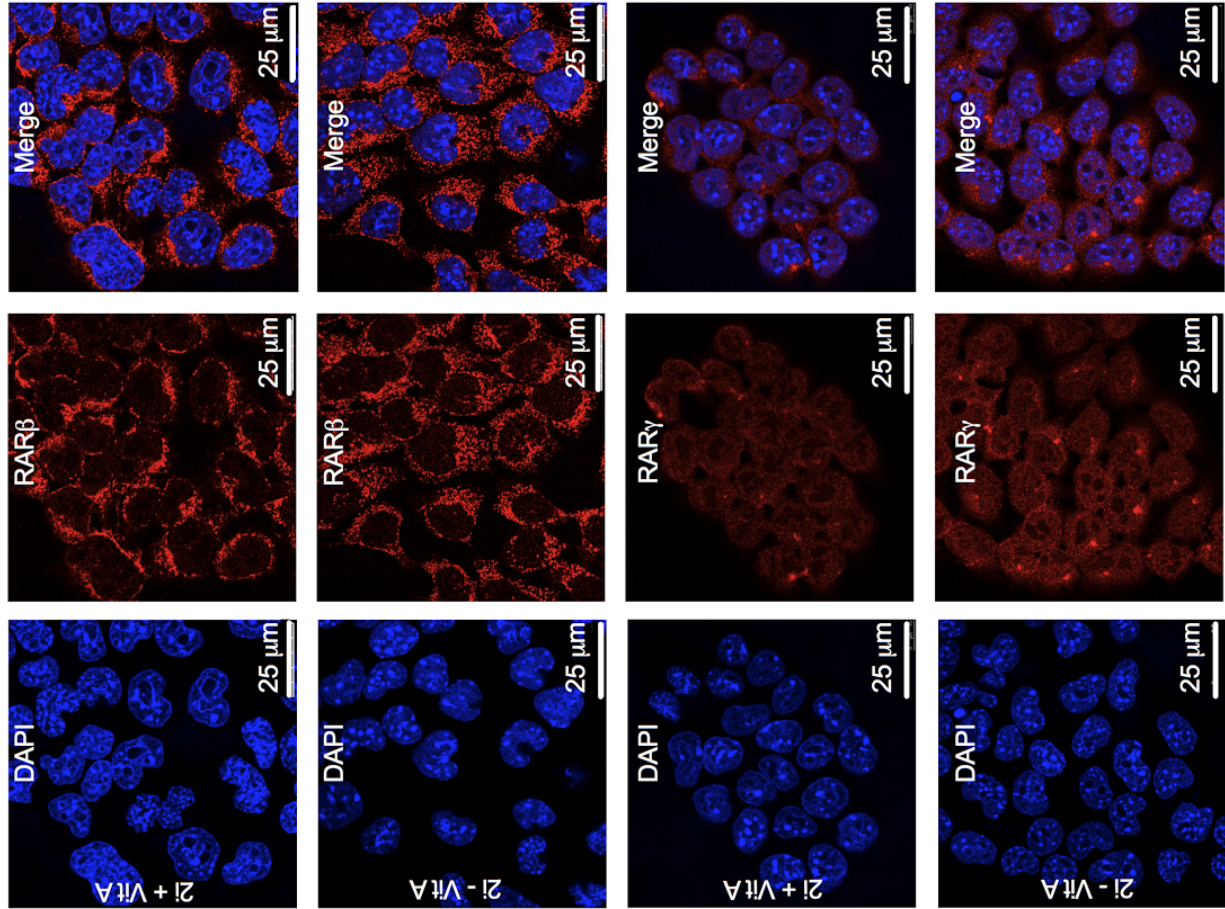
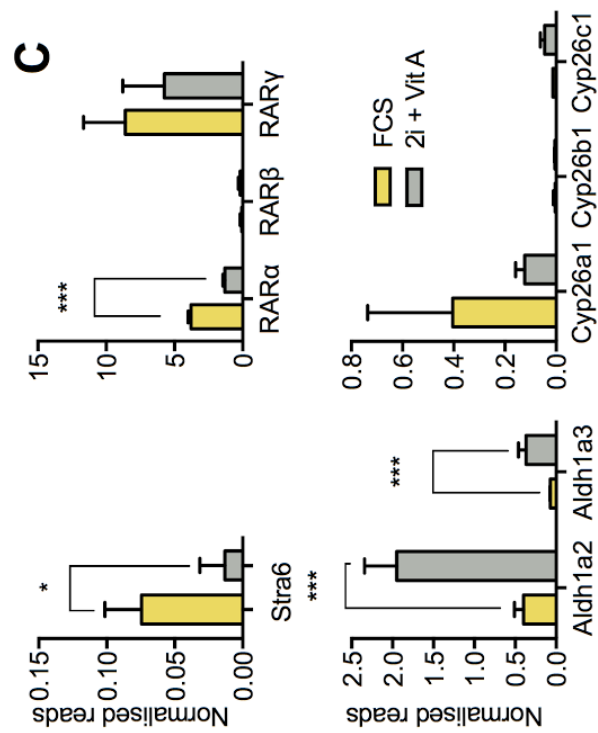
7.3.4. 2i *plus* Vitamin A activates PI3K and Akt signalling

It has been reported that retinol (vitamin A) is able to act as a bona fide growth factor via activation of the PI3K-Akt signalling pathway (Chen et al., 2010). Akt activation involves the phosphorylation of two residues: threonine 308 and serine 473. Thr308 is phosphorylated by the phosphoinositide dependent kinase PDK1 whilst Ser473 is phosphorylated by PDK2, a component of the mTOR complex 2. Once activated Akt translocates to various subcellular compartments, including the mitochondria, endoplasmic reticulum, the golgi apparatus and the nucleus. To assess whether the PI3K-Akt signalling pathway is more active in mESCs cultured with or without vitamin A, the phosphorylation of Thr308 and Ser473 was quantified by Western blot. These results were also confirmed by immunofluorescence staining, which enabled the cellular location of the phospho-Akt signal to be identified.

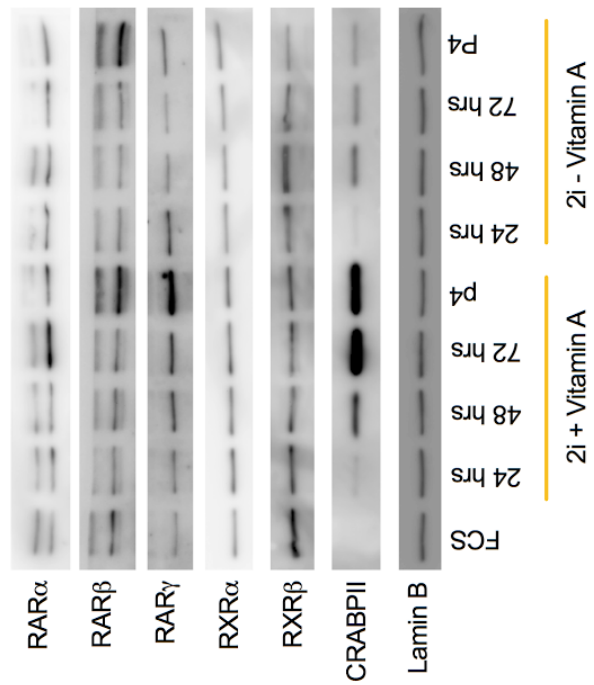
Fig. 7.6 Retinol (Vit A) does not activate the retinoic acid signalling pathway

(a) Expression of the ATRA transmembrane receptor and importer (Stra6), retinoic acid receptors (RAR), retinal dehydrogenases (Aldh), and ATRA metabolising enzymes (Cyp26) in mESCs cultured in FCS or 2i *with* vitamin A was quantified by RNA-Seq: * P-value < 0.05, ** P-value < 0.01, *** P-value < 0.0001. (b) Retinoic acid receptor expression was quantified by western blot. Lamin B was used as loading control. mESCs maintained in FCS were transferred to 2i *with* vitamin A or 2i *without* vitamin A and cell pellets were collected every 24 hours and after 4 passages in each condition. (c) mESCs maintained in 2i *with* vitamin A or 2i *without* vitamin A were stained with antibodies against the retinoic acid receptors RAR β and RAR γ .

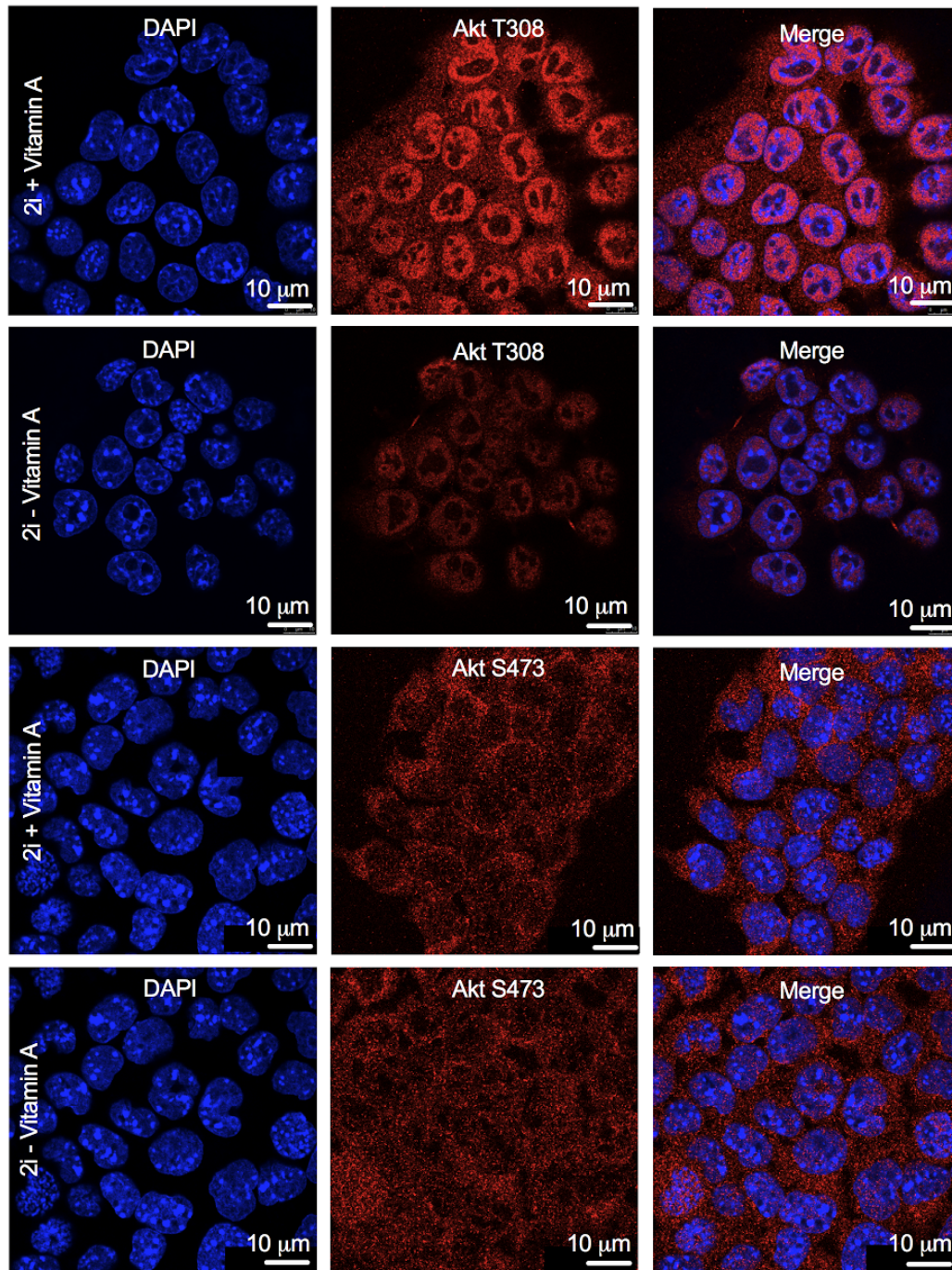
A



B



A



B

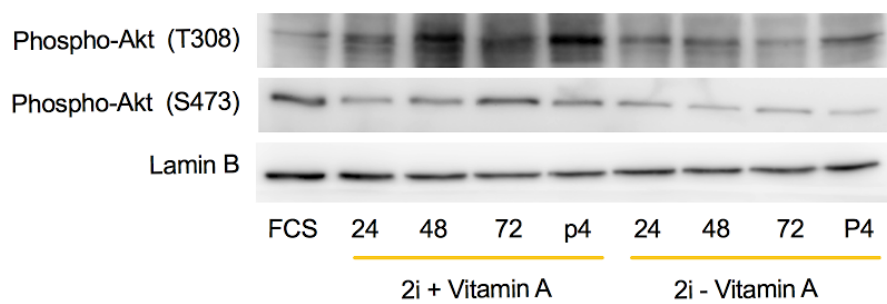


Fig 7.7 Akt signalling in the presence of vitamin A

(a) Immunofluorescent staining and cofocal microscopy of Akt activation in 2i *with* vitamin A and 2i *without* vitamin A. (b) WB analysis of Akt activation in mESCs maintained in 2i *with* vitamin A and 2i *without* vitamin A.

7.3.5. Cell cycle progression is dysregulated in 2i minus Vitamin A conditions

The constitutive activation of AKT leads to the dysregulation of the cell cycle and inhibition of apoptosis (Zhou et al., 2000). For example, AKT phosphorylates and inactivates GSK3 β . Given that GSK3 β typically phosphorylates Cyclin D and Myc, promoting their nuclear export and degradation, AKT phosphorylation and GSK3 β inactivation facilitates G1/S progression. Since that the expression of Dnmts does not substantially differ between 2i culture conditions with or without vitamin A and passive DNA demethylation is replication dependent, it is possible that the kinetics of DNA demethylation are driven by changes in the cell cycle. For instance, a shortened cell cycle may appear to show an increased rate of demethylation.

Cell cycle analysis of mESCs grown in FCS, 2i *plus* vitamin A and 2i *minus* vitamin A by dual EdU and DAPI staining confirmed that there are only minor differences in the proportion of cells residing in the early and late S phase when cells are grown in 2i *with* and *without* vitamin A. The profiles instead suggest that regardless of whether or not vitamin A is present, the cell cycle of mESCs grown in 2i differs substantially from those grown in FCS culture conditions in terms of the length of the G1 phase. mESCs are generally thought to have a short G1 phase, reflecting the bypassing of the restriction point. This is illustrated by mESCs grown in FCS conditions, where only 10% of cells are in G1. However, 20% and 25% of 2i *plus* vitamin A and 2i *minus* vitamin A cells respectively are found in G1 (7.8a).

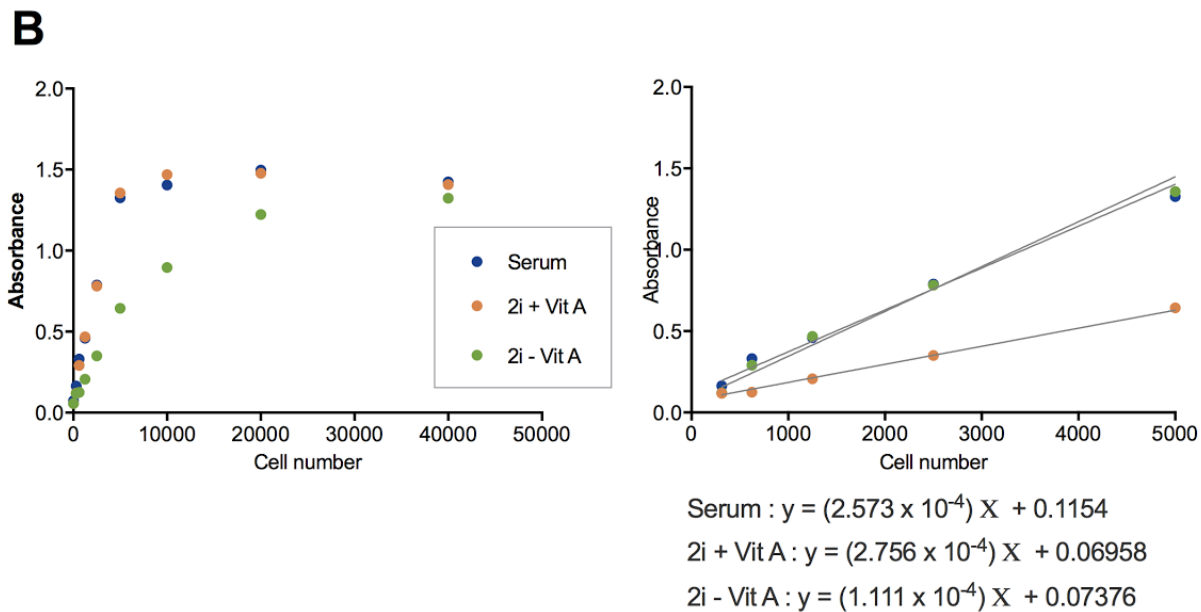
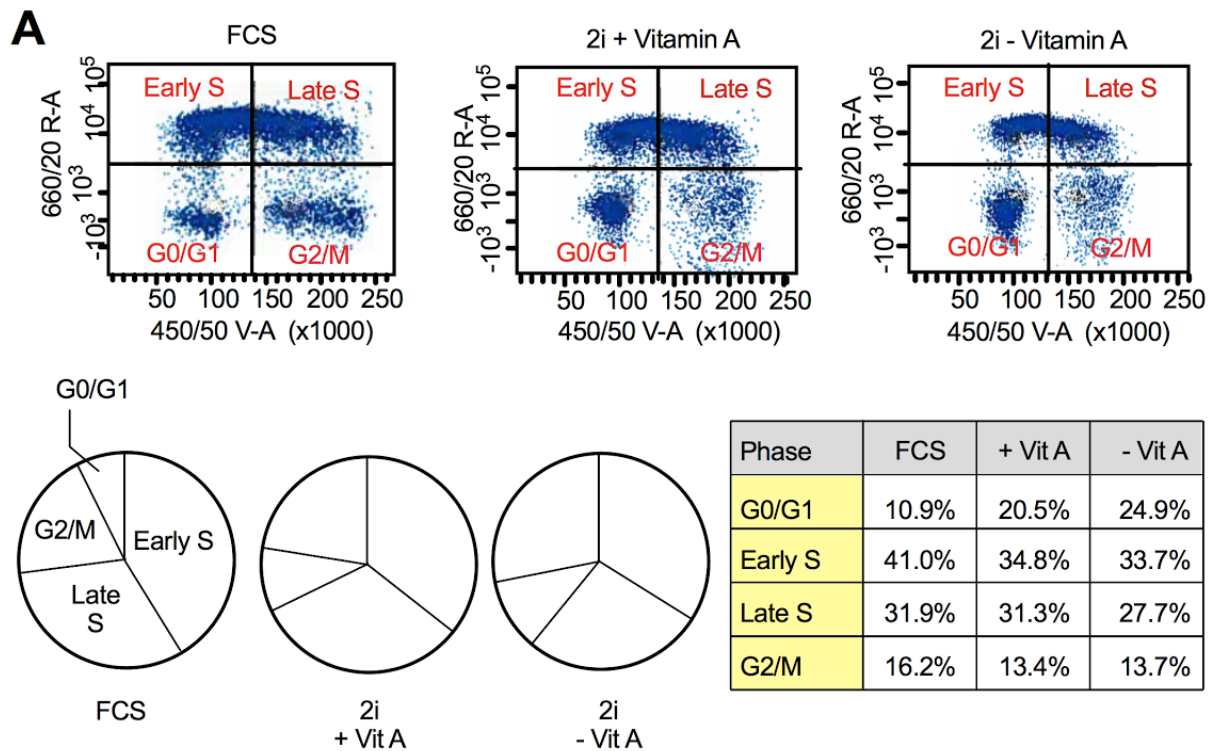


Figure 7.8 The absence of vitamin A causes defects in cell cycle progression

(a) mESCs (12g4) maintained in either FCS, 2i *with* vitamin A, or 2i *without* vitamin A were labelled with EdU for 15 minutes. EdU incorporation and DAPI staining were used to quantify the proportion of cells in either G0/G1, early S-phase, late S-phase/G2 or M-phase using flow cytometry.

(b) Serial dilutions of mESCs maintained in each of the three conditions were plated. The number of cells after 12 hours was determined by a colorimetric MTT assay.

Flow cytometric cell cycle analysis provides only a still of the number of cells in each cell cycle stage at any given time and therefore is not suitable for direct measurements of cell proliferation rate. Instead, proliferation dynamics were measured by counting subconfluent cells adapted in FCS and 2i. These results were confirmed using a colorimetric MTT cell proliferation assay and cell counting (Fig 7.8b).. Rates of cell proliferation were not substantially different in FCS or 2i. However, the presence of vitamin A on the overall length of the cell cycle was profound; the rate of cell proliferation without vitamin A was half that of cells grown with vitamin A, therefore adding substance to the hypothesis that Vitamin A enhances reprogramming to the naive ground state via activation of PI3K and Akt signalling and rapid cell proliferation.

7.3.6. Vitamin A and the DNMT1/UHRF1/PCNA complex

Alternatively, vitamin A may enhance replication-dependent DNA demethylation by reducing efficient maintenance methylation. Recruitment of DNMT1 to replication foci is facilitated by the interaction with UHRF1. Therefore, loss of UHRF1 or lack of interaction between DNMT1/UHRF1 and the replication foci could enhance DNA demethylation.

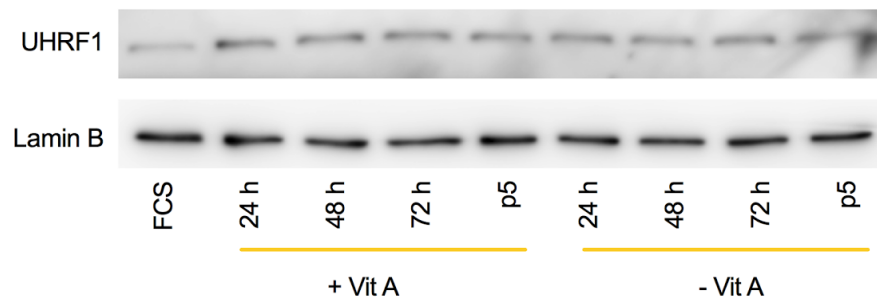


Figure 7.9 Culture-induced changes in Uhrf1 protein expression

12g4 maintained in FCS culture conditions were transferred to 2i condition. The expression of Uhrf1 was visualised by western blot during the first 72 hours after the addition of 2i and after 5 passages in 2i. Lamin B was used as a loading control.

No difference in UHRF1 expression could be detected by western blot, suggesting that this difference could not be explained by the downregulation of *Uhrf1*. Studies into the effect of post-translational modifications on DNMT1 activity are still in their infancy. However, it has been reported that these modifications can affect the

protein's stability and localisation (Scott et al., 2014)(Esteve et al., 2011)(Hodge et al., 2007). To assess whether the presence of vitamin A alters the co-localisation of DNMT1 with the replication fork, cells maintained in either FCS, 2i *plus* or *minus* Vitamin A conditions were labelled with EdU for 5 minutes and then co-stained with antibodies against either DNMT1, UHRF1, or both DNMT1 and UHRF1. Immunofluorescent labelling shows that DNMT1 colocalises with EdU foci in all three culture conditions. Therefore, passive DNA demethylation by inefficient DNMT1 recruitment is unlikely contribute to the more extensive DNA demethylation in 2i + vitamin A. To confirm this effect, it would be useful to investigate how vitamin A affects the interaction between DNMT1 and its protein partners. For instance, by immunoprecipitation of all DNMT1 interacting partners in each condition. Additionally, phosphorylation site mapping by mass-spectrometry would facilitate the identification of the sites responsible for the differential protein interactions in each condition.

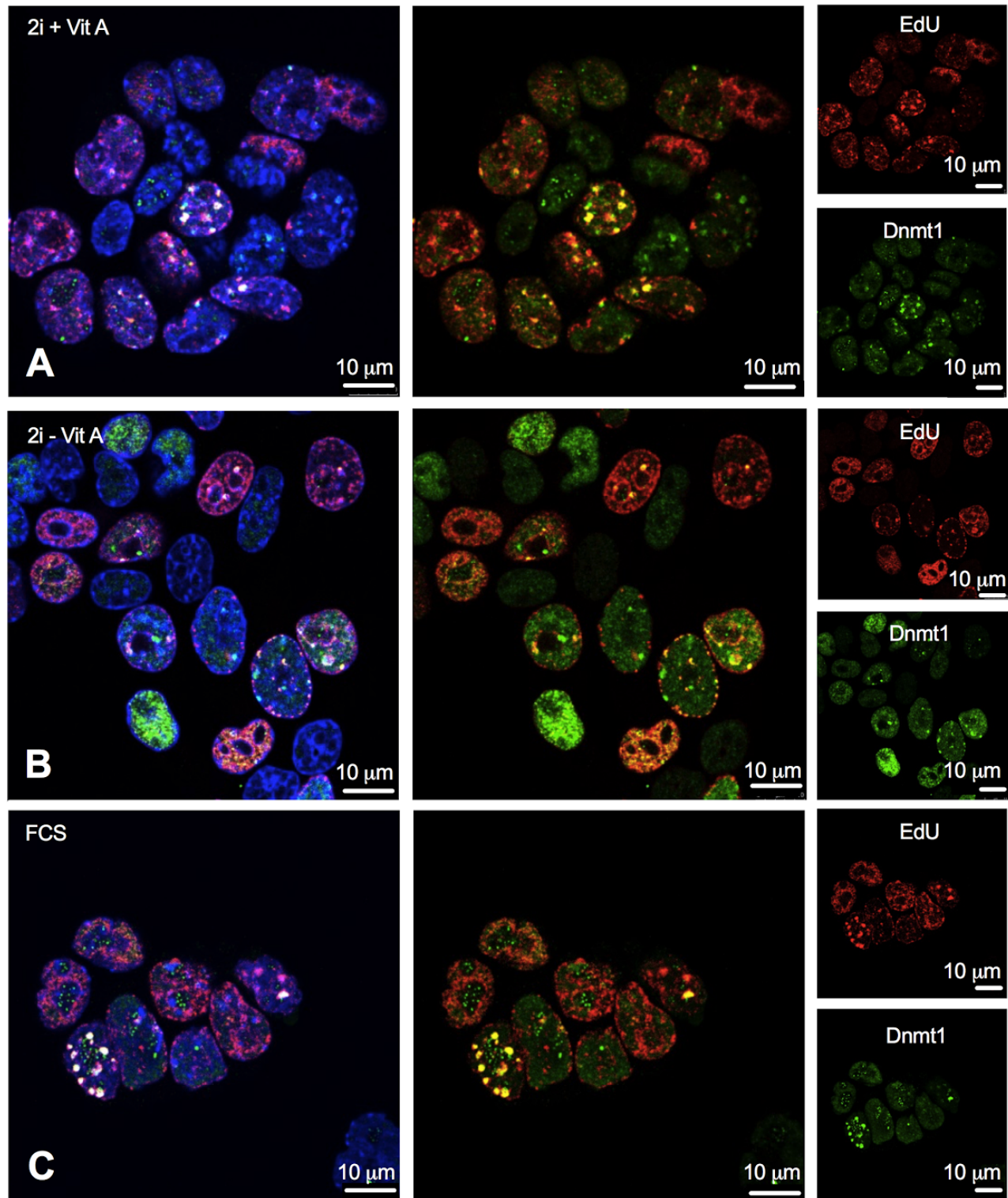


Figure 7.10 Colocalisation of Dnmt1 and EdU at replication foci

mESCs (12g4) maintained in either 2i *with* retinol (+Vit A) (a), 2i *without* retinol (-Vit A)(b), or FCS culture conditions (c) were labelled with EdU for 10 minutes. Incorporated EdU was covalently labelled with an Alexa fluor conjugated azide and cells were stained using an antibody against Dnmt1 (green). The localisation of Dnmt1 in comparison to the replication foci (depicted as localised regions of intense EdU labelling (red)), was visualised by confocal microscopy.

7.3 Discussion

Since the initial description of the 2i culture medium, many adaptations have been made to the original formulation, including the addition of vitamin C which is known to enhance DNA demethylation via the upregulation of TET activity (Blaschke et al., 2013). Other obvious differences include the amount of insulin used, the chosen substratum, and the presence or absence of vitamin A (retinol) in the B27 supplements. Additional differences are likely to exist, but are harder to define, particularly because many institutes prepare their own supplements. Whether commercial or made in house, it is often difficult to determine the exact formulation of supplements used in published works. Additionally, considerable lot-to-lot variation has been a concern for commercial N2 supplement, and components including transferrin and insulin are known to be denatured by freeze-thaw cycles. Therefore, subtle difference in cell culture conditions are likely to be found between 2i medium made in different laboratories. As the results presented in this chapter demonstrate, apparently nuance differences in culture medium can lead to very different experimental observations. Therefore, the use of standardised formulations and culture practices are a prerequisite if we are to draw accurate conclusions from the published data.

The work described in this chapter suggests that the presence of vitamin A is required for extensive 2i-induced DNA demethylation of mESCs. The mechanism behind this observation is not entirely obvious; the expression of both *de novo* and maintenance methyltransferases is not altered by the addition of vitamin A. Our initial findings suggest that vitamin A, rather than the morphogen all-trans retinoic acid, is the active metabolite responsible for the vast differences in global 5mdC. However, the vitamin A transporter Stra6 is transcriptionally downregulated in mESCs cultured in 2i conditions. If these changes reflect the downregulation of STRA6 at the protein level, channeling of vitamin A into the canonical retinoic acid signalling pathway seems unlikely. Furthermore, although the binding protein CRABP II is upregulated in 2i medium supplemented with B27 plus vitamin A, RARs are located, perhaps even sequestered, in distinct cytoplasmic foci. Under these conditions, it seems unlikely that RARs are able to act as transcriptional regulators. Instead, it has been

proposed that binding of vitamin A to cell surface receptors activates the Akt signalling pathway. Intracellular staining and western blotting is suggestive of increased phospho-Akt signalling.

Precisely how Akt signalling is altered in the presence of vitamin A remains an open question. The link between Akt signalling and DNA demethylation is also ambiguous. It is, however, appreciated that Akt signalling contributes to mESC self-renewal and the regulation of pluripotency (Wantanibe et al., 2006)(Pritsker et al., 2007). Consistently, we have observed that cells grown in 2i + vitamin A proliferate considerably faster compared to cells grown in 2i - vitamin A conditions. mESCs are thought to have a characteristically short G0/G1 phase, associated with the loss of MAPK, cyclin D, and pRb mediated control of cell cycle progression (Jirmanova et al., 2002)(Savaha., et al, 1996)(White et al., 2005). Differentiation is associated with an increase in the duration of G1, leading to suggestions that a shortened G1 phase is involved in the maintenance of the pluripotent state (Burdon et al., 2002)(Kapinas et al., 2013). Certainly, only 10% of mESCs cultured in FCS medium reside in G0/G1. However, a significantly higher proportion of mESCs cultured in 2i medium are in G0/G1, suggesting that this shortened phase is not a feature of ground state pluripotent cells. Interestingly, the S-phase is notably shorter in cells cultured without vitamin A. This was also obvious from our colocalisation studies, where there were very few cells labelled with EdU foci associated with late S/G2 phase.

Several reports have suggested that activated AKT signalling could result in the post translational modification of DNMT1. This area of research is still in its infancy. However, it is possible that post-translational modification of Dnmt1, particularly phosphorylation, could affect DNMT1/UHRF1/PCNA complex formation and therefore impact the efficiency of Dnmt1 recruitment to replication foci. Although dual labelling of Dnmt1 and EdU highlights that there are many more cells actively synthesising DNA in 2i + vitamin A culture conditions, compared to 2i - vitamin A, we do not observe a substantial fewer DNMT1-EdU foci in the presence of vitamin A. Therefore, to date we do not have evidence supporting the hypothesis that DNMT1 recruitment to replication foci is inefficient.

Assuming that *de novo* and maintenance methyltransferase activities are not altered by the presence of vitamin A, it seems likely that after prolonged passaging, the global level of 5mdC in cells maintained in each condition should equalise. Observations made by members of our group indicate that this is not the case. Instead DNA methylation levels plateau at approximately 75% their original value in cells maintained in FCS culture conditions (personal communication with Mr T. Huang, Imperial College London). It is possible that the vitamin A-dependent difference in DNA methylation state is indirect. The abundance of cofactors for DNA methylation and demethylating enzymes is potentially determined by the population's metabolic state. As such, it may be interesting to assess how different metabolic pathways are utilised in ground state pluripotent cells with and without vitamin A. The intersection between how metabolism is differentially regulated in pluripotent cells, and its role in defining the epigenetic landscape is a growing area of research and may shed some light on why this nuance has such a profound effect on 2i-triggered epigenetic reprogramming.

Chapter 8

Dissecting DNA demethylation and chromatin
dynamics during PGC reprogramming

8.1 Introduction

8.1.1. Pioneers of germ cell biology

The first comprehensive model of embryogenesis was presented in Aristotle's *De Generatione Animalium*. In essence, embryonic development was believed to be a gradual process, whereby organs sequentially form from a homogeneous starting material. Nowadays, this theory is known as epigenesis. In its earliest embodiment, epigenesis required a vital force (*vis essentialis*) to guide the developmental process. Following adoption by the catholic church, Aristotle's theory of epigenesis was highly influential throughout the Middle Ages. The Age of Enlightenment, however, saw a shift away from classical philosophies, in favour of materialism; the theory that processes are driven by physical matter, rather than intangible forces. In the study of embryonic development, materialism manifests as preformation; the idea that development is predetermined from conception and merely entails the growth or unfolding of pre-existing parts.

8.1.2. Primordial germ cell specification

Epigenesis and preformation theories have particular relevance to our understanding of germ cell specification. Model organisms, including *Xenopus*, *C. elegans* and *D. melanogaster* rely on the segregation of maternally derived cytoplasmic determinants (germ plasm) to specify nascent primordial germ cells (PGCs) and inhibit the somatic transcriptional programme. Accordingly, germ cells are defined shortly after fertilisation.

In contrast, mammalian germ cell development is *induced* by cell-cell signalling prior to the onset of gastrulation. Nascent PGCs are derived from the proximal epiblast. Nevertheless, isolated epiblasts are unable to support the specification of PGCs (Yoshimizu et al., 2001). PGCs can be identified based on high tissue non-specific alkaline phosphatase (TNAP) activity (Hahnel et al, 1990)(MacGregor et al., 1995) in the mouse embryo at E7.25, when a cluster of 40-50 nascent PGCs emerge in the proximal epiblast at the base of the allantois (Ginsburg et al., 1990)(Lawson et al., 1994). Transplantation studies have shown that any epiblast cell from the pre- or early

primitive streak stage embryo is capable of acquiring the germ cell fate (Tam et al., 1996). Therefore, the germline is defined by environmental cues rather than intrinsic features of proximal epiblast cells.

The secretion of BMP4 from the embryonic ectoderm is required for germ cell specification. Prior to the onset of gastrulation, BMP4 expression is confined to the extraembryonic ectoderm adjacent to the proximal epiblast. BMP4 null embryos fail to develop PGCs (Lawson, et al. 1999). BMP4 also activates Wnt3 (Ben-Haim et al., 2006), which is expressed in the proximal visceral endoderm (VE) and epiblast from E5.75 (Liu et al., 1999)(Rivera-Perez & Magnuson, 2005). The mesodermal factor Brachyury (T) is a key downstream effector of Wnt3. In prospective PGCs 'conditioned' by BMP4 signalling, T activates the transcription of the germ cell determinants Blimp1 (Prdm1), Prdm14, and AP2 γ (Ohinata et al., 2009)(Aramaki et al., 2013). Expression of this tripartite network is essential and sufficient for the silencing of the somatic transcriptional programme, the repression of the *de novo* DNA methyltransferases and the reactivation of a number of pluripotency factors (Magnusdottir et al, 2013)(Nakaki et al., 2014). These transcriptional changes are necessary to reverse mesodermal lineage restriction in favour of commitment to the germ line.

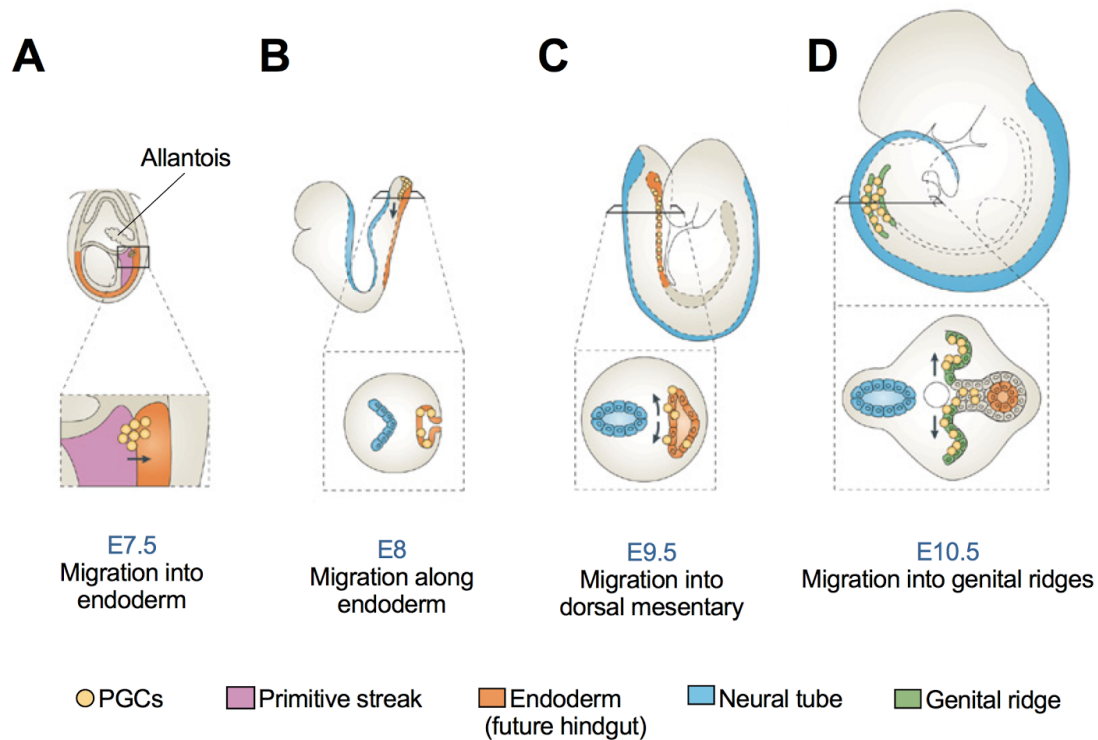


Figure 8.1 Stages of murine primordial germ cell migration

(a) Primordial germ cells (PGCs) are specified in the proximal epiblast at approximately E7.25 to E7.5. At E7.5, PGCs subsequently migrate from the primitive streak into the endoderm.

(b) Towards the end of gastrulation, at E8.0, the posterior visceral endoderm forms the hindgut, and distributes migratory PGCs along its length.

(c) At E9.5, PGCs actively migrate from the hindgut towards the dorsal body wall.

(d) By 10.5, PGCS colonise the genital ridge, forming the embryonic gonads.

Lateral views are shown on the top panel and transverse are shown on the bottom panel.

Adapted from Richardson & Lehmann (2010)

From a small cluster of cells at the midstreak stage, nascent PGCs migrate along the base of the allantois to the endoderm (Fig. 8.1a). Towards the end of gastrulation (approximately E8.0) the endoderm invaginates to produce the hindgut. Consequently, PGCs are dispersed along the length of the hindgut (Fig. 8.1b). From E9.5, PGCs migrate from the hindgut, through the dorsal mesentery and colonise the emerging genital ridges. By E11.5, all PGCs are gonadal (Fig. 8.1c-d).

8.1.3. Epigenetics of primordial germ cells

PGC development is accompanied by genome-wide epigenetic reprogramming, involving the erasure of DNA methylation, chromatin remodelling, and the transient loss of repressive histone modifications. Loss of DNA methylation (including the erasure of genomic imprints) is a prerequisite for the establishment of sex-specific imprints during gametogenesis, and the prevention of transgenerational epimutations. Moreover, germ cells must be able to create a totipotent zygote. However, prior to PGC specification, epiblast cell are primed for differentiation into the somatic germ layers. Reprogramming the epigenome is therefore a prerequisite to enable PGCs to enter a unique, pre-totipotent state.

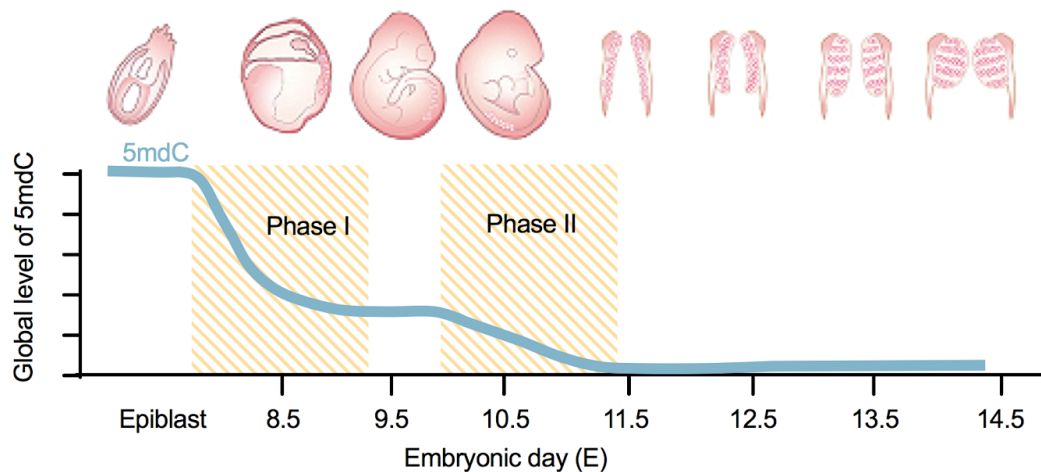


Figure 8.2 DNA demethylation dynamics during PGC development

DNA demethylation in developing PGCs occurs in two phases in migratory and gonadal PGCs respectively.

8.1.3.1 Migratory PGCs

Epigenetic reprogramming during PGC development is a highly orchestrated, multistep process. DNA demethylation occurs in two distinct waves in migratory and gonadal PGCs respectively (Fig. 8.2). Genome-wide methylation profiling has confirmed that migratory PGCs are hypomethylated compared to post-implantation epiblast cells (Seisenberger et al., 2012)(Guibert et al., 2012). DNA demethylation is likely to be a result of *de novo* DNMT downregulation (Kurimoto et al., 2008) and the passive, replication-dependent loss

of 5mdC (Seisenberger et al., 2012). Importantly, although the mRNA expression of *Dnmt1*, and its binding partner, *Uhrf1*, is slightly downregulated following PGC specification (Kurimoto et al., 2008), both proteins continue to be expressed throughout PGC development. Thus DNA demethylation occurs despite the continued presence of the maintenance DNMT.

DNA demethylation in migratory PGCs coincides with the reciprocal loss of H3K9me2 and increase in H3K27me3. The loss of H3K9me2 is likely to contribute to DNA demethylation. The GLP H3K9 methyltransferase is downregulated in nascent PGCs following specification (Hajkova et al., 2008)(Seki et al., 2007). In mESCs, loss of H3K9 methyltransferase activity is associated with reduced 5mdC at single copy genes, retrotransposons and major satellite repeats (Dong et al., 2008)(Tachibana et al., 2008). In somatic cells, the G9a H3K9 methyltransferase has been shown to interact directly with DNMT1 and to possibly coordinate the recruitment of DNMT1 to the replication foci. Therefore, reduced recruitment of DNMT1 may thus contribute to the passive loss of DNA methylation.

8.1.3.2 Gonadal PGCs

Following entry into the gonads at E11.5, PGCs undergo a second, more rapid wave of DNA demethylation (Guibert et al., 2012)(Hajkova et al., 2002)(Hajkova et al., 2008)(Hajkova et al., 2010). Based on the kinetics, DNA demethylation in gonadal PGCs is widely believed to be replication independent. Furthermore, stage 2 DNA demethylation is extensive. By E13.5 5mdC is erased from both genic and intergenic regions, imprinted domains and most repetitive elements (Hajkova et al., 2010)(Seisenberger et al., 2012)(Guibert et al., 2012)(Hajkova et al., 2002). As a result, global levels of 5mdC are lower in E12.5 PGCs than any other developmental stage.

The mechanism behind this second wave of DNA demethylation is unknown. However, the general consensus appears to favour an active mechanism. Genetic ablation of the deaminase *Aicda* has been shown to cause a slight hypermethylation of PGCs (Popp et al., 2010). Notably, *Aicda* knockout does not cause embryonic lethality or loss of fertility, and therefore these results are controversial. Alternatively, DNA demethylation in

non-migratory PGCs coincides with the upregulation of Tet1 and Tet2 (Hajkova et al., 2010). The resulting 5hmdC may be removed by further oxidation, excision and BER DNA repair or replication-dependent dilution (Hackett et al., 2013)(Kagiwada et al, 2013).

Studies by our group have shown that DNA demethylation during PGC development is not impaired in the *Tet1* knockout mice (personal communication with Dr P. Hill). Global quantification of DNA modifications by LC-MS has revealed that 5hmdC abundance is not elevated following DNA demethylation in gonadal PGCs. Therefore, the global DNA demethylation in gonadal PGCs does not appear to be linked to increased TET activity. Furthermore, Tet1 KO PGCs are not hypermethylated compared to wild type PGCs on the same background (Hill et al., unpublished). Thus further experiments are required to shed light on the mechanisms underlying DNA demethylation in gonadal PGCs.

Epigenetic reprogramming in the gonadal PGCs also entails the reactivation of the inactive X-chromosome (Xi) (Hajkova et al., 2002)(Monk & McLaren, 1981)(Tam et al, 1994)(de Napoles et al, 2007) and the transient loss of repressive histone modifications, including H3K9me3, H3K27me3, loss of the linker histone H1 and widespread chromatin decondensation (Hajkova et al, 2008). This is typically seen as an enlargement of the nuclei. The widespread loss of histone modifications and chromatin decondensation suggest that reprogramming is connected to large scale histone replacement. Accordingly, the histone chaperones NAP1 and HIRA are highly expressed in PGCs compared to neighbouring somatic cells (Hajkova et al., 2008).

8.1.3.3. The repair model of epigenetic reprogramming

In gonadal PGCs, DNA demethylation precedes the activation of the base excision DNA repair (BER) pathway; as shown by an increase in PARylation and the recruitment of the scaffold DNA repair protein XRCC1 to chromatin. The BER pathway evolved as a means of protecting cells from endogenous DNA damage induced by hydrolysis, reactive oxygen species and intracellular metabolites that chemically modify the DNA bases. PARP1 is an abundant nuclear protein that constitutively associates with chromatin in a predominantly inactive state. DNA single strand breaks (SSBs), such as those formed by the excision of an abasic nucleotide, activate Parp1.

Studies on *Drosophila* polytene chromosomes have shown that there is a strong correlation between ADP-ribose accumulation and the decondensation of chromatin. These studies suggest that polyADP-ribosylation (PARylation) potentially leads to nucleosome disassembly and histone replacement. According to the Repair Model of epigenetic reprogramming, the activation of the BER DNA repair pathway following TET-driven 5mdC hydroxylation may induce widespread histone replacement and chromatin changes (Hajkova et al., 2008).

8.1.3.4 *De novo* methylation in germ cells

Following DNA demethylation, female germ cells initiate meiosis at E13.5 and arrest in diplotene of prophase I. Primary oocytes are incorporated into primary follicles after birth. The growth of an activated (primary) follicle into a secondary (early antral) follicle is accompanied by *de novo* methylation. To the contrary, male germ cells, arrest in mitotic G1 at E13.5, prior to entry into meiosis and *de novo* DNA methylation occurs later on in the foetal testes (Henckel et al., 2009)(Davis et al., 2000). Failure to re-establish appropriate methylation results in the reactivation of repeat elements, meiotic defects and failed gametogenesis. The male germ line is sustained by a population of spermatogonial stem cells. Given their reliance on multiple rounds of cell division, the faithful maintenance of DNA methylation is particularly important for male gonocytes.

8.2 Experimental aims

Epigenetic reprogramming in gonadal PGCs appears to be highly complex and the mechanisms behind both DNA demethylation and widespread chromatin remodelling are areas of active research. I was particularly interested in how these two processes are mechanistically linked or coordinated. Therefore, my aim was to develop an experimental system which could potentially uncouple the loss of DNA methylation and histone replacement. Using LC-MS and immunofluorescence confocal microscopy, I wanted to quantify the changes in abundance of DNA modifications and both histone modifications and BER DNA repair proteins respectively, to assess whether these processes are inextricably linked.

Epigenetic reprogramming is associated with widespread changes in the transcriptional programme; from the somatic programme of the late epiblast and the restricted reactivation of the pluripotency network in nascent PGCs to the preparation for meiosis in gonadal PCGs. Therefore, I also wanted to understand how these switches in gene expression are triggered and how corresponding changes in the epigenetic landscape affect the process.

8.3 Results

8.3.1 Experimental system

Dnmt1, is the sole DNA methyltransferase expressed throughout PGC development (Hajkova et al., 2002)(Seki et al., 2005)(Yabuta et al., 2006). Genetic ablation should therefore lead to the passive, replication-dependent dilution of 5mdC. Deletion of *Dnmt1* in mice results in embryonic lethality prior to midgestation (Li, Bestor & Jaenisch, 1992). Therefore, we wanted to develop a conditional, germline *Dnmt1* knockout model. In the absence of *de novo* DNA methyltransferases, *Dnmt1* deletion will induce rapid, replication dependent DNA demethylation in proliferating migratory PGCs. Therefore, we expected *Dnmt1* KO PGCs to be hypomethylated compared to wild type PGCs of the same developmental stage. By artificially demethylating the PGC genome before entry into the embryonic gonads, we hoped to remove the substrate for active DNA demethylation and the potential mechanistic link between DNA demethylation and chromatin remodeling.

The timing of *Dnmt1* deletion is critical. Several different promoters can be used to specifically drive the cre-mediated germline deletion of *Dnmt1*, including *Blimp1* (*Prdm1*), *Tnap* and *Stra8*. Although the expression of *Tnap* is detectable at E7.25 (Ginsburg et al., 1980)(Hahnel et al., 1990)(MacGregor et al., 1995), *Tnap-Cre* driven deletion is inefficient and unlikely to result in the absence of DNMT1 protein by the onset of gonadal epigenetic reprogramming (Lomeli et al., 1999). *Stra8* on the other hand is expressed in female gonadal PGCs at approximately E12.5 (Menke et al., 2003), and is therefore also unsuitable for the deletion of *Dnmt1* in early PGCs. *Prdm1* expression can be identified in PGC precursors from as early as E6.25 (Lawson & Hage, 1994)(Ohinata et al., 2005)(Vincent et al., 2005). However, the only available *Prdm1-Cre* transgenic line is known to have a severe developmental phenotype; significantly fewer pups are born compared to the average C57B6 litter size. This appears to be due to embryonic lethality at mid-gestation. Nonetheless, *Prdm1* is currently the only suitable driver of *Dnmt1* deletion and was therefore used for all the *in vivo* experiments described in this chapter.

We obtained *Dnmt1* *F/F* mice (B6.129S4-*Dnmt1*^{tm2^{jae}}) from Jackson Laboratories. *Prdm1-Cre* was specifically used to drive Cre-Lox recombination between LoxP sites flanking *Dnmt1* exons 4 and 5; creating an out of frame splice from exon 3 to exon 6 (Fig. 8.3a). The resulting RNA encodes only the first 67 amino acids and lacks the localization motifs and the entire catalytic domain. Additionally, homozygous *Dnmt1* *F/F* mice were crossed with GGOF mice carrying a Δ *PE-Oct4 GFP* transgene. GFP expression, driven by the *Oct4* distal enhancer, is specific to naive mESCs and PGCs and can facilitate the isolation of PGCs from the surrounding soma for further downstream analysis. A schematic of the breeding scheme is shown in figure 8.3b.

8.3.2 *Dnmt1* KO PGCs fail to develop into mature male germ cells

Testes isolated from 4 week old conditional *Dnmt1* *F/+* (wild type) and *Dnmt1* Δ/Δ *Prdm1-Cre* (KO) littermates showed a substantial difference in size (Fig. 8.4a). The total mass of the mice did not differ significantly. Haematoxylin and eosin (H&E) staining from frozen tissue sections shows a complete lack of spermatogenesis in *Dnmt1* Δ/Δ *Prdm1-Cre* testes (Fig. 8.4c). *Dnmt1* *F/+* testes however, clearly show a layer of spermatogonial cells adjacent to the basal lamina of the seminiferous tubule (identified by their round nucleus), followed by spermatocytes and spermatids in increasingly luminal positions. Mature spermatozoa can be seen in neither the *Dnmt1* *F/+* nor *Dnmt1* Δ/Δ *Prdm1-Cre* testes. This is likely due to the use of prepubescent males.

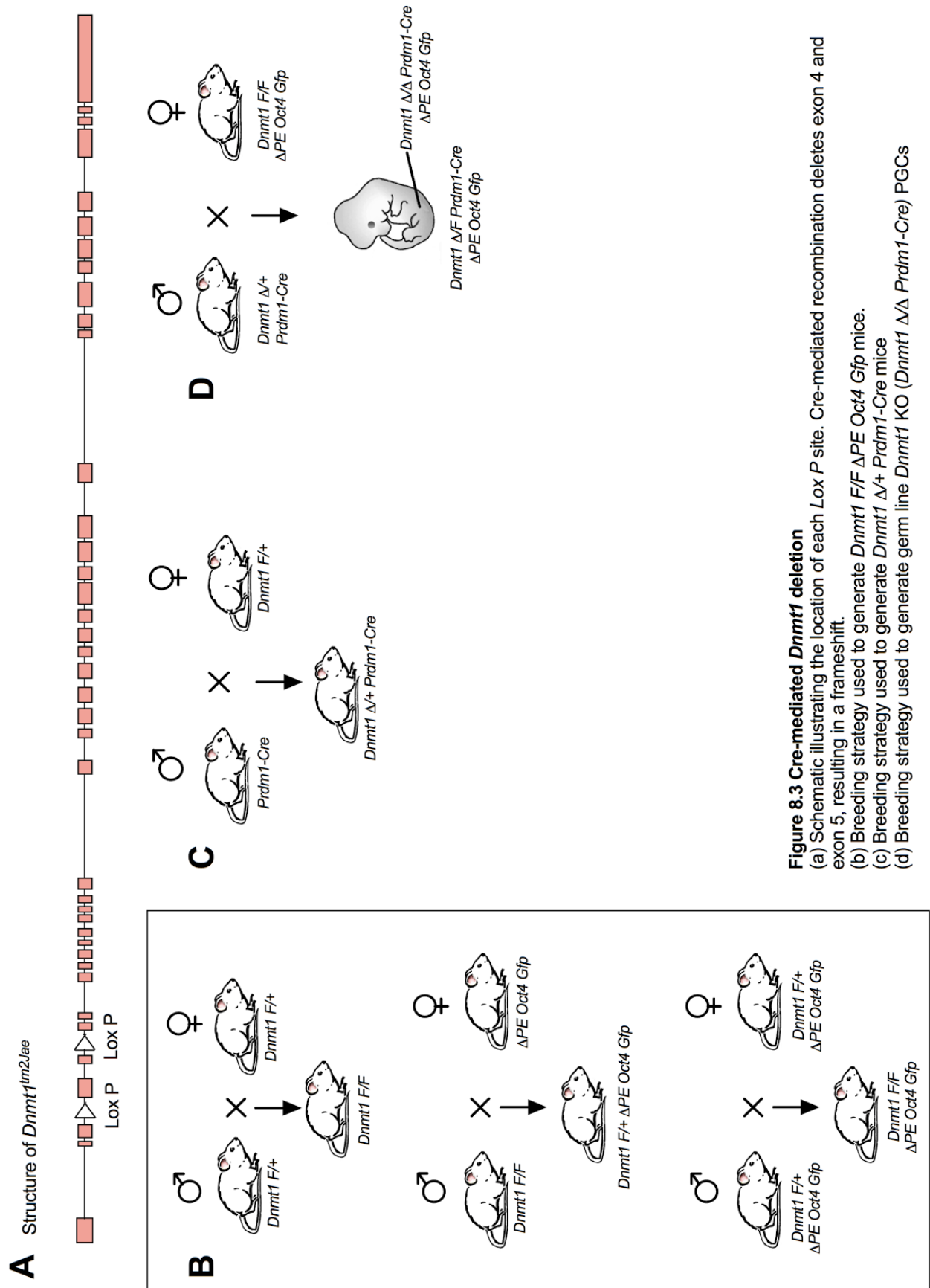


Figure 8.3 Cre-mediated *Dnmt1* deletion
 (a) Schematic illustrating the location of each Lox P site. Cre-mediated recombination deletes exon 4 and exon 5, resulting in a frameshift.
 (b) Breeding strategy used to generate *Dnmt1 F/F ΔPE Oct4 Gfp* mice.
 (c) Breeding strategy used to generate *Dnmt1 Δ/+ Prdm1-Cre* mice
 (d) Breeding strategy used to generate germ line *Dnmt1* KO (*Dnmt1 Δ/Δ Prdm1-Cre*) PGCs

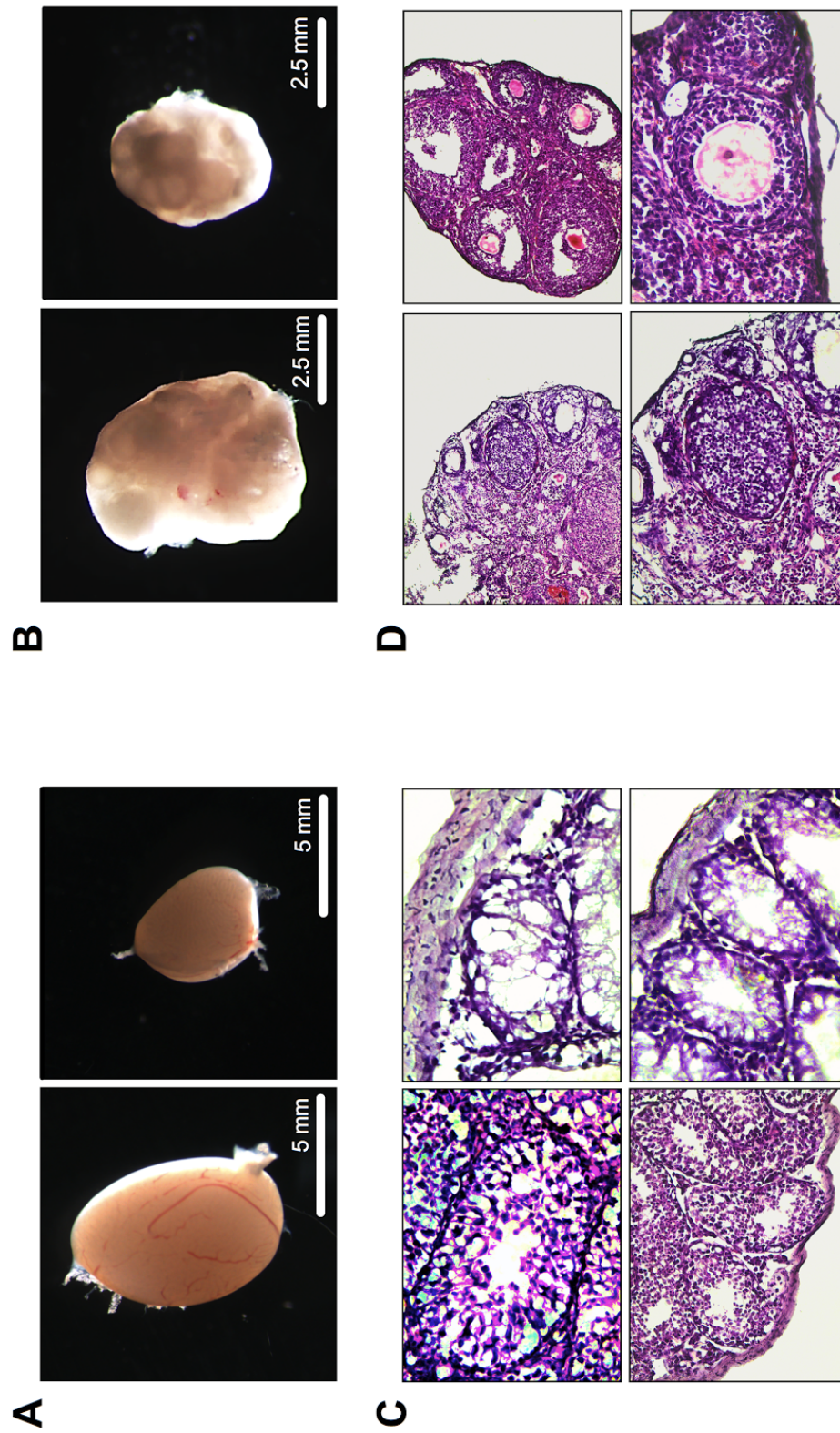


Figure 8.4 Phenotype of germline *Dnmt1* deletion in adult gonads
(a) *Dnmt1* *f/+* and *Dnmt1* Δ/Δ *Prdm1-Cre* testis and ovaries were removed from 4 week old males and 6 week old females respectively. The *Dnmt1* *F/Δ* *Prdm1-Cre* and *Dnmt1* *f/+* mice were from the same litter and did not differ significantly in size or weight. The gonads were fixed in 4% PFA for 48 hours and frozen in liquid nitrogen. (b) 10 μ m sections were cut and stained with haematoxylin and eosin to visualise the gross tissue structure.

Similarly the size of ovaries dissected from 6 week old *Dnmt1 F/+* and *Dnmt1 Δ/Δ Prdm1-Cre* ovaries are remarkably different (Fig. 8.4b). H&E staining of frozen sections did, however, show that at least some oocytes were able to enter follicle development in the *Dnmt1 Δ/Δ Prdm1-Cre* ovary (Fig. 8.4d). These observations may reflect the differences in gametogenesis following sex differentiation. Spermatogenesis relies on early *de novo* DNA methylation and the maintenance of DNA methylation states following multiple rounds of DNA synthesis whilst in females *de novo* DNA methylation occurs in the non-proliferative, activated follicle and is dependent on DNMT3A and DNMT3L. The obvious difference in size between *Dnmt1 F/+* and *Dnmt1 Δ/Δ Prdm1-Cre* ovaries may nonetheless indicate that deletion of *Dnmt1* is detrimental to oocyte development. It will be interesting to assess how *Dnmt1* affects the number of primary follicles and whether *Dnmt1 Δ/Δ Prdm1-Cre* oocytes are able to produce viable offspring.

8.3.3 DNMT1 is depleted prior to entry into the genital ridges

Following *Prdm1-Cre* driven recombination, DNMT1 protein is depleted in migratory PGCs by E10.5 (Fig. 8.5). To confirm that the loss of methyltransferase activity translates into extensive global DNA demethylation, we must now quantify the level of DNA modifications in developing *Dnmt1* KO and wild type PGCs. Using the LC-MS methods described in previous chapters, we will quantify the global levels of DNA modifications throughout PGC development to determine whether PGC genomes are demethylated before entry into the genital ridge. Following the loss of DNMT1, germ cell numbers (detected using anti-DDX4) are severely depleted in the male embryonic gonads (Fig. 8.6). Since *Dnmt1 Δ/Δ Prdm1-Cre* male and female adult gonads have such different phenotypes, it will be important to determine whether *Dnmt1 Δ/Δ Prdm1-Cre* female germ cells are also depleted in E14.5 embryonic gonads.

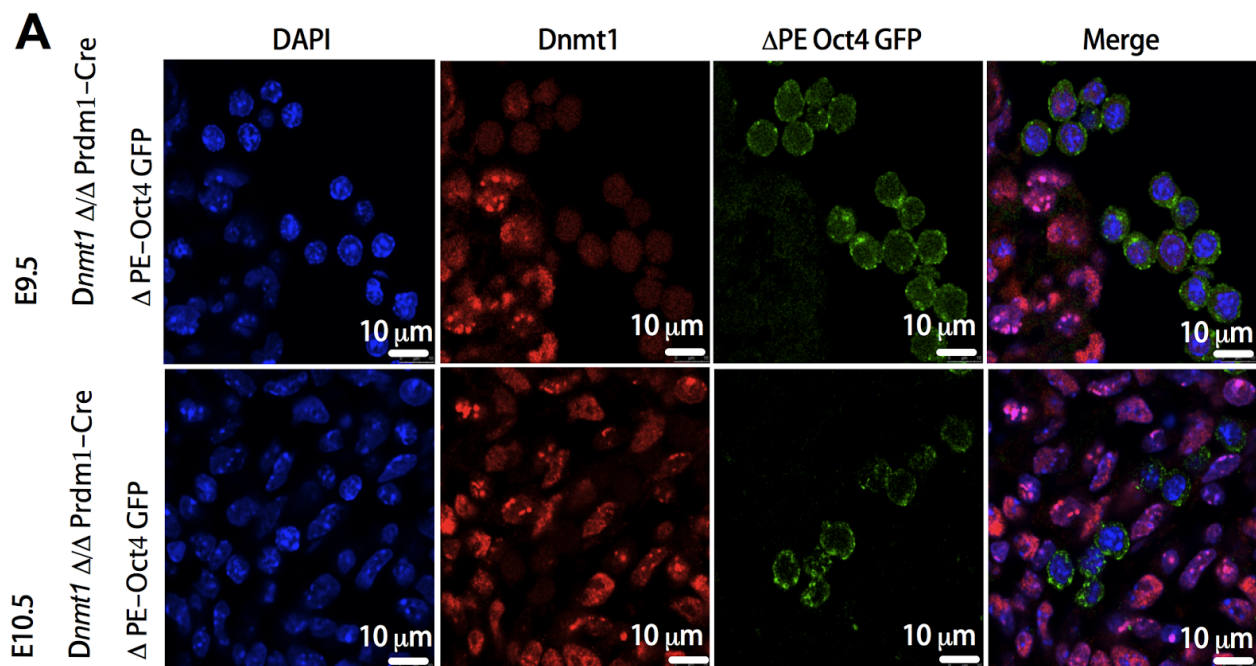


Figure 8.5 DNMT1 is depleted in E10.5 PGCs

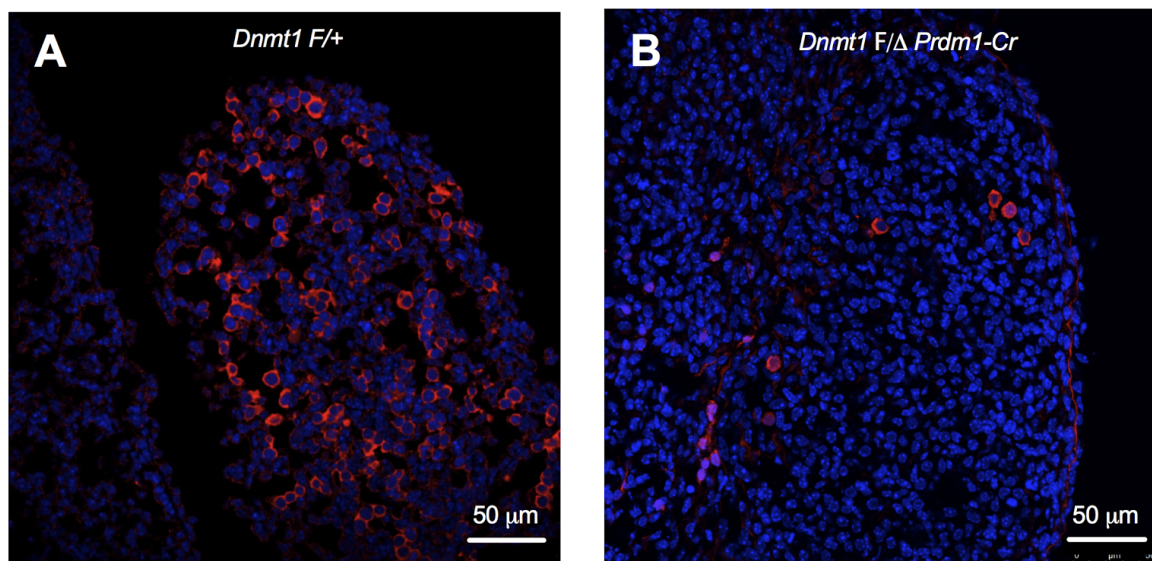


Figure 8.6 Number of germ cells in male E14.5 gonads

Gonads were isolated from E14.5 *Dnmt1* F/+ (a) and *Dnmt1* F/ Δ Prdm1-Cr (b) embryos. The number of germ cells was visualised using anti-DDX4/MVH (red). DNA is stained with DAPI (blue).

8.3.4. Developing *ex vivo* models of germline *Dnmt1* knockout

PGC development is a unique physiological example of *in vivo* epigenetic reprogramming. However, PGCs can only be obtained in low numbers and these studies are not compatible with many commonly used biological techniques. Alternative, *ex vivo* models are a potentially useful means of complementing *in vivo* studies of epigenetic reprogramming within the post-implantation embryo. Recently, viable, fertile and healthy offspring have been produced from PGC-like cells (PGCLCs) induced from naive mESCs (Hayashi et al., 2012). Simplistically, this model relies on a stepwise differentiation from mESC to primed epiblast-like cells, followed by the induction of PGCLC differentiation using same molecular signals produced by the extraembryonic ectoderm to specify nascent PGCs (Fig. 8.7a). This system therefore appeared to be an ideally suited to explore the relationship between active DNA demethylation and chromatin remodelling dynamics in the germ line.

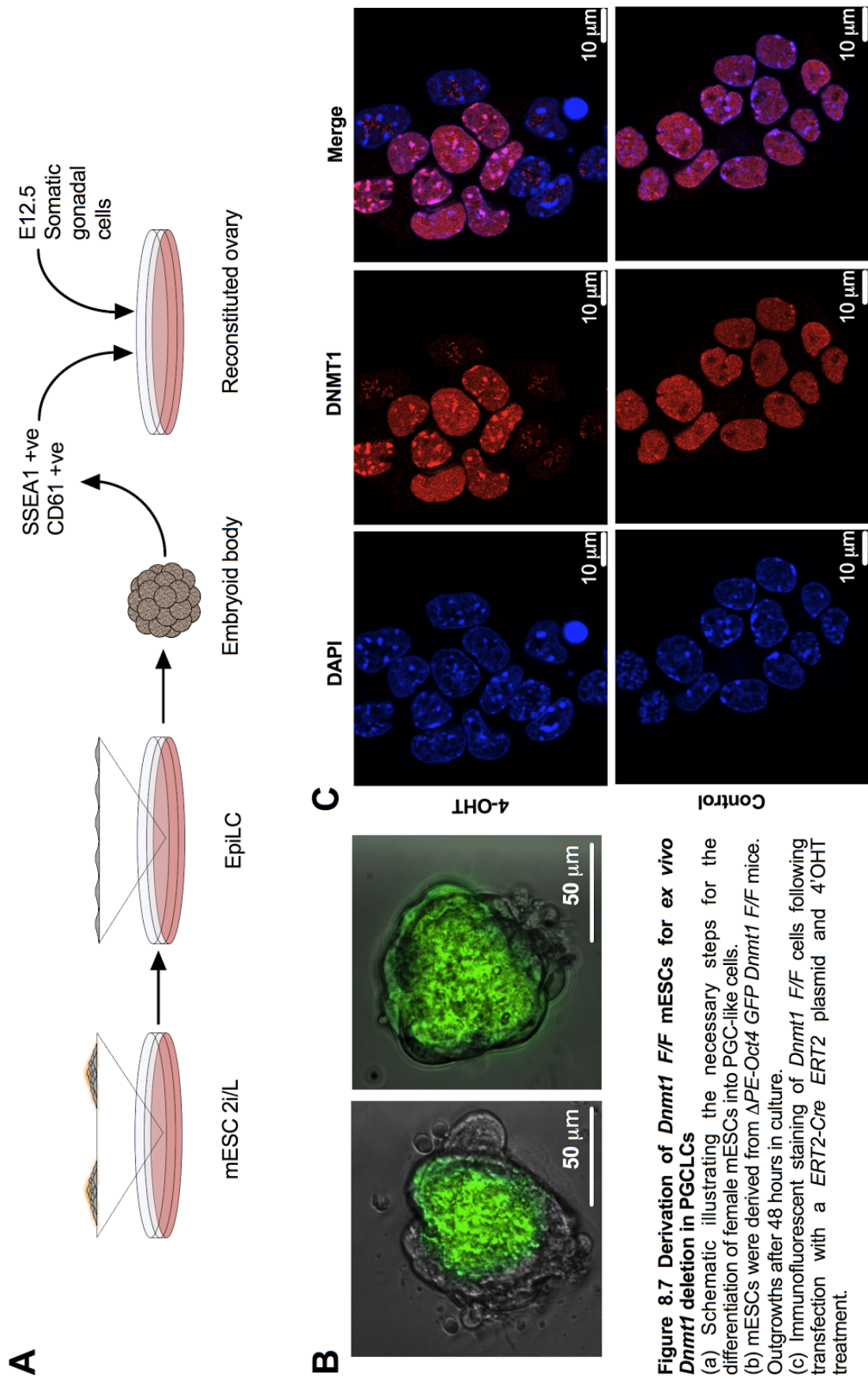


Figure 8.7 Derivation of *Dnmt1* F/F mESCs for ex vivo *Dnmt1* deletion in PGCLCs

(a) Schematic illustrating the necessary steps for the differentiation of female mESCs into PGC-like cells.

(b) mESCs were derived from Δ PE-Oct4 GFP *Dnmt1* F/F mice. Outgrowths after 48 hours in culture.

(c) Immunofluorescent staining of *Dnmt1* F/F cells following transfection with a ERT2-Cre ERT2 plasmid and 4'OHT treatment.

8.3.5. Derivation and characterisation of *Dnmt1* f/f mESCs

DNMT1 is required for the transition from mESCs to EpiLCs. Therefore, in order to delete *Dnmt1* during the course of PGC development, *Dnmt1* deletion had to be inducible; for instance by using a tamoxifen inducible Cre. With the technical expertise of Dr. H. Leitch, I derived homozygous *Dnmt1* f/f Δ PE-Oct4 GFP mESC lines. Briefly, blastocysts were collected at E3.5 and cultured in KSOM until a discernable blastocoel could be seen. Typically, the ICM is isolated by immunosurgery; acid Tyrode's solution disintegrates the zona pellucida whilst complement-mediated lysis removes the surrounding trophoctoderm cells. However, we found that degradation of the zona pellucida (hatching) occurs spontaneously in KSOM and therefore immunosurgery is unnecessary. Single late-stage blastocysts were cultured in 2i on a layer of MEFs until a clear pluripotent outgrowth could be seen (Fig. 8.7b). The outgrowths were subsequently enzymatically disaggregated and passaged in FCS-free and feeder-free conditions.

In total, 16 mESC lines were expanded. Each line was genotyped to confirm that the cells were of *Dnmt1* f/f Δ PE-Oct4 GFP genotype. Additionally, the sex of each cell line was determined using primers for both the sex-determining region of the Y chromosome (*Sry*) and *SmcY/SmcX*. All cell lines identified as female were further characterised for the expression of pluripotency markers. Robust *Oct4* expression was determined based on Δ PE-Oct4 GFP signal intensity and the expression of NANOG, OCT4 and SOX2 protein was assessed using immunofluorescence labelling and confocal microscopy. Additionally, criteria including typical naive mESC morphology (compact, refractive colonies of clonogenic cells) was used to select a subset of female mESC lines to use for future experiments.

Selected *Dnmt1* f/f Δ PE-Oct4 GFP mESC lines were transfected with a lentiviral vector carrying a CMV-promoter driven *ERT2-Cre ERT2* (Cre-recombinase oestrogen T2 receptor) transgene. Cells carrying the linearised plasmid were selected based on the expression a puromycin resistance cassette. Cells from individual puromycin-resistant colonies were isolated and maintained in medium supplemented with puromycin for a

further passage. No further cell death was seen, indicating that all cells contained the transgene. Additionally, each mESC line was genotyped using universal *Cre* primers.

A subset of clones were treated with $0.25 \mu\text{g ml}^{-1}$ 4' hydroxytamoxifen (4'OHT) for up to 96 hours. Cre-mediated recombination should be a rapid event, however genotyping for *Dnmt1* exon 4 revealed that a *Dnmt1 f/f* allele was present in all samples. To understand why the deletion was not effective, I quantified the expression of *Cre* mRNA by RT-qPCR. Cre expression was not detected in any of the clones, but was detected in a control mESC line generated by another member of our group. This confirmed that the transgene was silenced due to *de novo* DNA methylation of the CMV promoter.

The Rosa26 locus is frequently used as a site of transgene insertion due to the relative ease of homologous recombination and the ubiquitous and constitutive expression of transgenes from the endogenous Rosa promoter, without transcriptional silencing. Following electroporation with a plasmid carrying *ERT2-Cre ERT2* within the Rosa26 locus, multiple clones were isolated and screened for the presence of Cre transgene mRNA. In total, 3 separate *Dnmt1 f/f* Δ *PE-Oct4 GFP* mESC lines were transfected and 25 puromycin resistant colonies were picked from each cell line. Based on RT-qPCR quantification, clones with high expression of Cre were chosen for tamoxifen-induced LoxP recombination. As an initial screen, the 4 clones from each cell line expressing the highest levels of Cre mRNA were treated with $1.0 \mu\text{g ml}^{-1}$ 4'OHT for 72 hours. The cells were then split, with half being collected for DNA extraction whilst the remainder were plated on coverslips for immunofluorescence labelling of DNMT1 protein.

Genotyping for the presence of *Dnmt1* exon 4 indicated that once again, recombination had not occurred in a proportion of cells in each sample. This could be due to a) mutations within at least one of the *LoxP* sites preventing recombination, b) deletion of exons 4 and 5 only a single allele in each cell or c) a mosaic pattern of *Dnmt1* deletion. Both *LoxP* sites were sequenced, and found to contain no mutations. Immunofluorescence labelling and confocal microscopy identified that, whilst *Dnmt1* is successfully deleted in the majority of the cell population, high levels of DNMT1 protein are found in a small fraction of cells. Repeating the tamoxifen

treatment with the addition of puromycin produces the same mosaic result (Fig. 8.7c); suggesting that contamination with untransfected cells is not responsible for this observation. Assuming therefore that each clone is derived from a single transfection event, it is difficult to explain this mosaic deletion.

8.4 Discussion

The results presented in this chapter show the development of an experimental system to explore the mechanisms driving DNA demethylation and chromatin remodelling during PGC development. I have shown that, when driven by the *Prdm1* promoter, *Dnmt1* is deleted in developing PGCs by embryonic day (E)9.5; prior to the second phase of DNA demethylation. Since the *in vivo* model described in this chapter relies generating a genomic template devoid of 5mdC *before* the proposed onset of active DNA demethylation, it is now necessary to quantify the kinetics of 5mdC loss following *Dnmt1* deletion. Our LC-MS methods have previously been used to quantify the abundance of DNA modifications in PGCs lacking functional TET1. The same approach would be used to create an accurate and detailed depiction of replication-dependent DNA demethylation following loss of the maintenance methyltransferase.

The effects of precocious DNA demethylation are likely to be widespread. Many germline specific genes are controlled by CGI-linked promoters. These regulatory elements are constitutively methylated in somatic cells, but become hypomethylated during DNA demethylation in PGCs, coinciding with their transcriptional activation. Typically, these genes encode regulators of meiosis, including *Dazl* and *Sycp3*. Manipulating DNA demethylation dynamics in developing PGCs has the potential to cause widespread changes to the PGC transcriptome. These differences are unlikely to be fully appreciated without deep sequencing.

Finally, to determine whether active DNA demethylation via TET-mediated oxidation and chromatin remodelling are mechanistically linked by activation of the BER repair pathway, it would be important to quantify the changes in chromatin remodelling following *Dnmt1* deletion, including the loss of linker histone

H1, the deposition of H2A/H4R3me2s, the localisation of histone chaperones including Nap1, and the deposition and removal of the histone variants H2A.X and H2A.Z, and activation of the BER DNA repair pathway.

Primordial germ cells are ultimately unipotent and under normal developmental conditions are only capable of differentiating into gametes. However, ectopic transplantation of midgestational genital ridges triggers teratocarcinoma formation (Stevens et al., 1984) and *ex vivo* PGCs can give rise to pluripotent embryonic germ (EG) cells (Matsui et al., 1992)(Resnick et al., 1992). Pluripotent EG cells share many fundamental properties with ES cells, including morphology, a stable, normal karyotype and the ability to contribute to chimaeras (Labosky et al., 1994)(Stewart, 1994). Additionally, dual MEK and GSK3 β inhibition in 2i conditions leads to comparable transcriptome changes and global DNA hypomethylation in both cell types (Leitch et al., 2013). Understandably, the shared characteristics of ES and EG cells has led to comparisons of their respective cells of origin; the naive epiblast and PGCs (Surani & Reik, 2007). For instance, the pluripotency transcriptional network is known to be upregulated following PGC specification, and both the early epiblast and PGCs are hypomethylated compared to the post-implantation epiblast and the surrounding soma.

The capacity to tolerate global loss of DNA methylation appears to be specific to pluripotent stem cells. *Dnmt1* null mESCs stably maintain global DNA methylation levels with approximately 20% 5mdC found in wildtype mESCs. These cells proliferate normally without the apparent loss of pluripotency (Tsumura et al., 2006)(Li, 1996)(Panning & Jaenisch, 1996). Furthermore, *Dnmt1*, *Dnmt3a* and *Dnmt3b* triple knockout (TKO) mESCs contain no detectable 5mdC but can be maintained in an undifferentiated state. Notably, both *Dnmt1* KO and *Dnmt* TKO mESCs show pronounced differentiation defects. Interestingly, hESCs are generally thought of as 'primed', and cannot tolerate the loss of *Dnmt1*; suggesting that the capacity to withstand DNA hypomethylation is a property of naive pluripotency rather than pluripotency *per se*.

Dnmt1 deletion is known to trigger apoptosis in a multitude of different cell types, including haematopoietic progenitor cells (Trowbridge et al., 2009), adult epidermal stem cells (Sen et al., 2010), pancreatic progenitor

cells (Georgia et al., 2013), the HCT116 human colorectal cancer cell line (Chen et al., 2007) and cultured spermatogonial stem cells (germline stem cells, GC cells)(Takashima et al., 2013). Apoptosis appears to be preceded by G2/M cell cycle arrest and, following release from this state of arrest, is triggered by p53 during mitosis (Chen et al., 2007). Cell death triggered by *Dnmt1* deletion can be circumvented by the additional deletion of p53 (Takashima et al, 2013).

Although we have yet to confirm precisely when PGCs numbers begin to decline, the contrasting phenotypes between *Dnmt1* null male and female adult gonads suggest that apoptosis occurs after the divergence of the sex-specific developmental pathways. This deleterious effect of *Dnmt1* deletion may be exacerbated in male gonocytes due to their high rates of proliferation compared to female gonocytes. Unlike ovaries, which contain a finite number of primary oocytes at the time of birth, spermatogonial stem cells continually generate spermatogonia throughout the adult male's lifetime. It is conceivable that loss of maintenance methyltransferase activity in proliferating cells is incompatible with spermatogonial development.

However, severe loss of male PGCs is apparent already at E14.5, and therefore cannot be entirely explained by differences in gonocyte proliferation. Additionally, since DNA methylation in male germ cells is reacquired from E15.5 onwards, widespread PGC death cannot be attributed to the failure to establish male-germ cell specific methylation patterns. Apoptosis may instead be caused by the premature erasure of parent-of-origin imprints, the reactivation of retrotransposons, or the precocious expression of germline specific genes from CGI-linked promoters and the dysregulated preparation for meiosis.

Since beginning this project, our group has made several key discoveries that cast doubts over the widely accepted model of active TET-dependent DNA demethylation in the mouse germline. Specifically, DNA demethylation in gonadal PGCs is largely independent of TET activity, with only a subset of loci being affected by the lack of 5hmC formation. Several models have been proposed to explain these observations. Hydroxylation may instead be important for the maintenance of the hypomethylated state; removing aberrant *de novo* DNA methylation, similarly to DNA demethylation in the zygote (Amouroux et al., 2016). Alternatively,

hydroxymethylation may be required to specifically remove 5mdC from particular loci to enable transcriptional activation. Finally, since TET proteins directly interact with chromatin modifiers, TET-mediated hydroxylation may simply be a side-effect of TET recruitment alongside other chromatin modifying activities.

Increasingly, it seems unlikely that TET-mediated hydroxylation is *directly* mechanistically linked to the genome wide activation of BER-pathway in PGCs. Furthermore, BER DNA repair is activated in *Tet* TKO zygotes. The physical association between TET, chromatin remodelling factors, and proteins involved in BER DNA repair may instead cause the off-target hydroxylation of 5mdC at sites targeted for nucleosome replacement. In this instance, we may expect to see activation of BER, despite not having a template for active DNA demethylation. Alternatively, the BER pathway may be active at only a subset of genomic loci, where TET-driven hydroxylation and chromatin remodelling is required for transcriptional activation.

It is however clear that PGC development is a highly orchestrated process. Given the potential for PGCs to form teratocarcinomas, and the importance of the correct execution of the germline transcriptional programme for the fitness of the next generation, it is highly likely that fail-safe mechanisms are in place to halt the development of PGCs that deviate from this well trodden path. This system could also be useful to further our understanding of how the timing of DNA methylation, chromatin remodelling, activation of the germline transcriptional programme and progression into meiosis are related, and the consequences of dysregulation.

Conclusion

DNA methylation patterns are associated with the stabilisation of cell fate. Under certain conditions, including the *ex vivo* pharmacological inhibition of signaling pathways and the *in vivo* epigenetic reprogramming of primordial germ cells, DNA methylation states can be remodelled. The molecular processes that underpin genome-wide epigenetic reprogramming are far from understood. Despite extensive research, many fundamental questions remain unanswered. In part, *in vivo* studies of epigenetic reprogramming in PGCs are severely hampered by the technical, and sometimes ethical, challenges associated with working with cells of limited availability. Our group has invested significant time in developing a highly sensitive LC-MS method for the quantification of DNA modifications in small numbers of cells. To date, this method has enabled us to conclude that both DNA demethylation in the zygote and gonadal PGCs is not directly linked to TET-mediated hydroxylation (Amouroux et al., 2016)(Hill et al., in preparation). .

There are many parallels between 2i-induced epigenetic reprogramming in migratory PGCs and epigenetic reprogramming associated with the transition of mESCs from FCS-based growth medium to 2i medium. Hence, the aforementioned experimental mESC culture system could be used as an alternative and complementary model to study DNA demethylation in migratory PGCs. Using *Dicer KO* mESCs, I have identified that functional miRNA processing is necessary for 2i-induced DNA demethylation, and more specifically for the post-transcriptional downregulation of ESC-specific *Dnmt3a isoform 2*, in mESCs. It has previously been suggested that upregulation of the transcriptional repressor *Prdm14* precedes and is ultimately responsible for the downregulation of the *de novo* DNA methyltransferases. However, based on the kinetics of DNA methyltransferase downregulation and *Prdm14* upregulation, this seems unlikely. Rather, I suggest that the initial loss of DNA methylation may instead be triggered by miRNA-mediated regulation of the *de novo* DNA methyltransferases; the upregulation of *Prdm14* being a later event necessary for the stable repression of the *de novo* DNA methyltransferases (Fig. 9.1). I further suggest that widespread changes in the miRNA transcriptome are required to reprogramme mESCs to the pluripotent

ground state. These changes may redefine the signaling landscape; for example, by setting the threshold over which a cell will respond to particular stimuli.

The idea that DNA methylation may be regulated by miRNAs is not novel. The binding of specific miRNAs to DNA methyltransferase transcripts has been experimentally validated. miRNAs are also known to control de novo DNA methylation by the regulation of transcriptional repressors (Sinkkonen et al., 2008). Although a number of studies show that miRNAs regulate germ cell development (Banisch et al., 2012), their role in specifically regulating DNA demethylation in migratory PGCs has not yet been explored. Therefore, it may be interesting to look for parallels between miRNA expression in mESCs cultured in 2i and miRNAs abundantly expressed in migratory PGCs.

Of course, if true, the implications of this work would extend beyond merely furthering our understanding of pluripotency and germ cell biology. An increasing number of molecular therapies rely on the production of induced pluripotent stem cells (iPSCs) and gene editing. miRNAs are already known to facilitate somatic cell reprogramming (Anokye-Danso et al., 2011) and thus a set of miRNAs specifically involved in restoring naive pluripotency may be used to generate naive iPSCs. The phenotype of naive iPSCs, including clonogenicity, high rates of proliferation, and the ability to produce cross-species chimeras are all highly desirable for the development of these therapies. Furthermore, the transcriptome analysis described in Chapter 6 highlights several similarities between *Dicer* KO mESCs and human ESCs (hESCs); including the upregulation of the epithelial-mesenchymal transition (EMT) and the expression of XIST. The generation of ground state hESCs is technically challenging (Takashima et al., 2014). miRNAs specifically expressed in ground state mESCs may facilitate or enhance the efficiency hESC reprogramming.

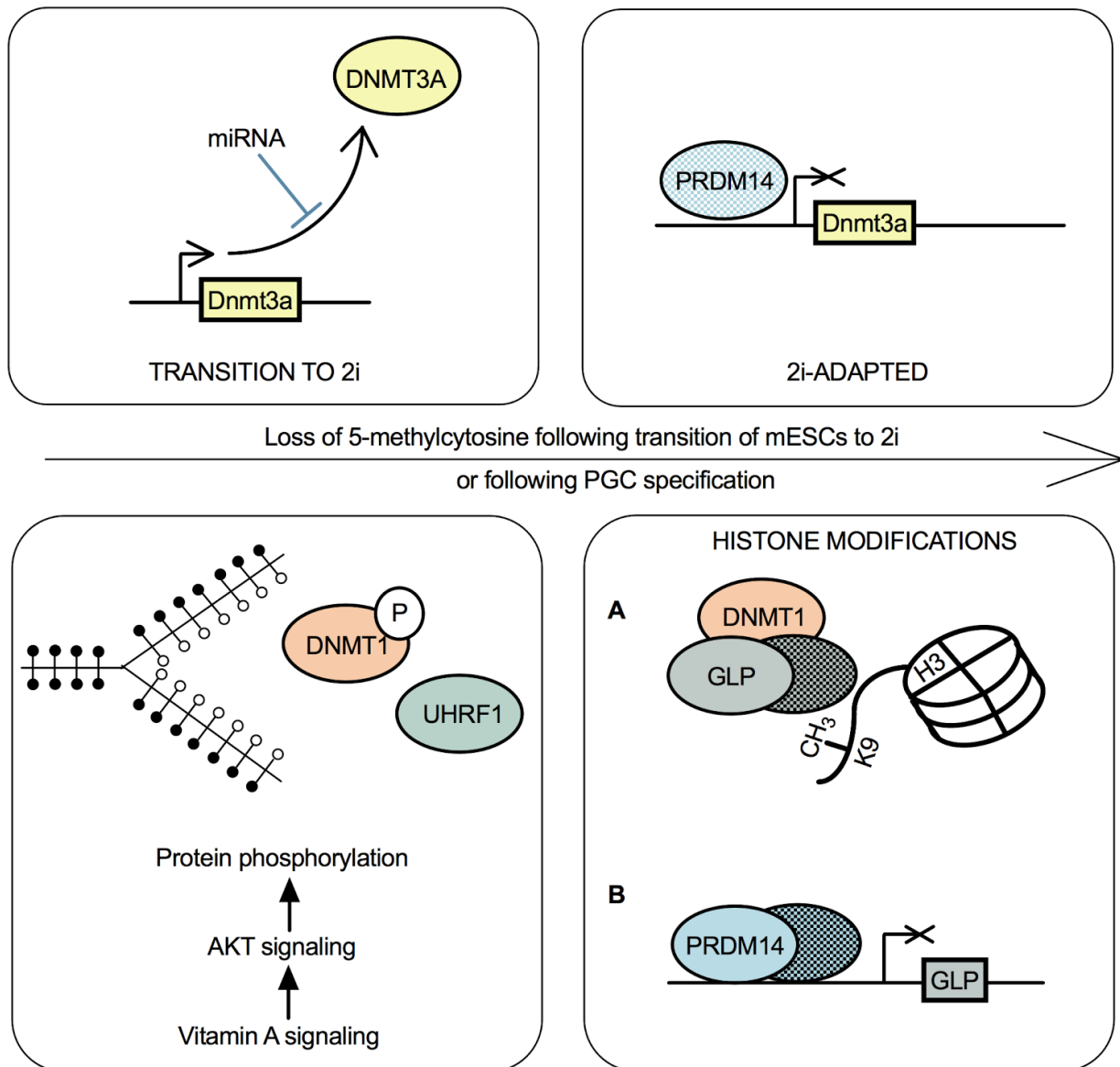


Figure 9.1: Mechanisms suggested to contribute to the global loss of 5-methylcytosine during the reprogramming of mESCs to the naive ground state and during the reprogramming of migratory primordial germ cells.

As a word of caution, I have also demonstrated that seemingly small variations in culture conditions can profoundly affect DNA demethylation. In particular, I found that vitamin A is required for rapid and extensive DNA demethylation in mESCs upon transition to 2i medium. On one level, this observation may go some way to explain several inconsistencies in the published literature. How vitamin A facilitates the loss of DNA methylation is not clear. However, I have shown that vitamin A is unlikely to act via the canonical retinoic acid signaling pathway.

This is perhaps unsurprising given the potent differentiation-inducing effect of retinoic acid. My preliminary results instead suggest that vitamin A activates the AKT signaling pathway (Fig. 9.1); a pathway known to regulate “stemness” in a variety of stem cell systems (Watanabe et al., 2006). Further, AKT signaling has been shown to promote the derivation of embryonic germ (EG) cells from PGCs (Kimura et al., 2008). Whether increased AKT signalling is an immediate effect of vitamin A or simply a downstream consequence of heightened pluripotency is not yet clear.

Although not the focus of my studies, it is inevitable that histone modifications and chromatin remodelling also contribute to DNA methylation changes in mESCs and migratory PGCs. Since these regulatory tiers cannot be neatly resolved it is likely that my model is highly simplified. For instance, I have considered PRDM14 as primarily a transcriptional repressor of the *de novo* DNA methyltransferases. It is known however, that PRDM14 is also a transcriptional repressor of the histone methyltransferase GLP (Yamaji et al., 2008). GLP and the related molecule G9a modulate gene repression via the methylation of lysine 9 of histone H3. Accordingly, loss of H3K9me2 is characteristic of epigenetic reprogramming in both mESCs and migratory PGCs. GLP/G9a are also known to interact directly with DNMT1, thereby coordinating transcriptional repression by both DNA methylation and histone modifications (Fig 9.1). Therefore, it is possible that PRDM14 may primarily modulate global histone modifications; the DNA methylation changes merely being secondary effects.

My model of DNA demethylation in mESCs and migratory PGCs is further complicated by the idea that, although classically considered a stable epigenetic modification, 5-methylcytosine may actually be in a constant state of flux; being largely driven by DNA replication, and also by DNA repair and transcription. By combining LC-MS and isotope tracing studies, I have developed a system to analyse the dynamic, steady state of DNA methylation using *in vivo* and *ex vivo* models. I show that 5-methylcytosine in postmitotic adult cortical neurons can be labelled using $^{13}\text{C-d}_3$ L-methionine; thereby providing evidence for the turnover of 5-methylcytosine in the context of a post-replicative cell. If our preliminary results are confirmed, the implications of this finding could be widespread. It would be particularly important to understand how DNA methylation patterns are faithfully maintained in a state of constant flux, and whether dynamic DNA methylation can contribute to cell adaptability and

developmental plasticity. It is my hope that this system can also be utilised to dissect the kinetics of sequential 5mdC modifications during periods of epigenetic reprogramming.

In all, I have described a series of experimental systems and technologies for use in dissecting the process of DNA demethylation in the mouse germline. More broadly, these methods may be used to answer the fundamental question of how a somatic cell from the post-implantation epiblast can be reconfigured to enable the perpetuation of genetic and epigenetic information between generations. This field has suffered from major technical challenges; with studies purporting to provide evidence for multiple, often contradictory, mechanistic pathways. Several of the methods described herein have already been used to debunk widely held beliefs relating to this process. It is my hope that with the aid of more reliable and accurate technologies, the research community will continue to generate more concrete conclusions.

Bibliography

Aapola, Ulla et al. "Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22. 3, related to the cytosine-5-methyltransferase 3 gene family." *Genomics* 65.3 (2000): 293-298.

Adelman, Carrie A, Subrata Chattopadhyay, and James J Bieker. "The BMP/BMPR/Smad pathway directs expression of the erythroid-specific EKLF and GATA1 transcription factors during embryoid body differentiation in serum-free media." *Development* 129.2 (2002): 539-549.

Amouroux, Rachel et al. "De novo DNA methylation drives 5hmC accumulation in mouse zygotes." *Nature cell biology* 18.2 (2016): 225-233.

Annesley, Thomas M. "Ion suppression in mass spectrometry." *Clinical chemistry* 49.7 (2003): 1041-1044.

Aramaki, Shinya et al. "A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants." *Developmental cell* 27.5 (2013): 516-529.

Arand, Julia et al. "In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases." *PLoS Genet* 8.6 (2012): e1002750.

Arita, Kyohei et al. "Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism." *Nature* 455.7214 (2008): 818-821.

Artzt, Karen et al. "Surface antigens common to mouse cleavage embryos and primitive teratocarcinoma cells in culture." *Proceedings of the National Academy of Sciences* 70.10 (1973): 2988-2992.

Arvey, Aaron et al. "Sequence and chromatin determinants of cell-type-specific transcription factor binding." *Genome research* 22.9 (2012): 1723-1734.

Avantaggiato, Virginia et al. "Retinoic acid induces stage-specific repatterning of the rostral central nervous system." *Developmental biology* 175.2 (1996): 347-357.

Avvakumov, George V et al. "Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1." *Nature* 455.7214 (2008): 822-825.

Bachman, Martin et al. "5-Hydroxymethylcytosine is a predominantly stable DNA modification." *Nature chemistry* 6.12 (2014): 1049-1055.

Bachman, Martin et al. "5-Formylcytosine can be a stable DNA modification in mammals." *Nature chemical biology* 11.8 (2015): 555-557.

Bader, Scott et al. "Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency." *Oncogene* 18.56 (1999).

Bain, Jenny et al. "The selectivity of protein kinase inhibitors: a further update." *Biochemical Journal* 408.3 (2007): 297-315.

Barreto, Guillermo et al. "Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation." *Nature* 445.7128 (2007): 671-675.

Bellacosa, Alfonso et al. "MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1." *Proceedings of the National Academy of Sciences* 96.7 (1999): 3969-3974.

Bernstein, Emily, Sang Yong Kim, Michelle A. Carmell, Elizabeth P. Murchison, Heather Alcorn, Mamie Z. Li, Alea A. Mills, Stephen J. Elledge, Kathryn V. Anderson, and Gregory J. Hannon. "Dicer Is Essential for Mouse Development." *Nature Genetics* 35.3 (2003): 215-17.

Bestor, Timothy H, and Vernon M Ingram. "Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA." *Proceedings of the National Academy of Sciences* 80.18 (1983): 5559-5563.

Bhatia, Himanshu et al. "Maintenance of feeder free anchorage independent cultures of ES and iPS cells by retinol/vitamin A." *Journal of cellular biochemistry* 113.9 (2012): 3002-3010.

Bhattacharya, Sanjoy K et al. "A mammalian protein with specific demethylase activity for mCpG DNA." *Nature* 397.6720 (1999): 579-583.

Bhutani, Nidhi et al. "Reprogramming towards pluripotency requires AID-dependent DNA demethylation." *Nature* 463.7284 (2010): 1042-1047.

Birchler, James A, Manika Pal Bhadra, and Utpal Bhadra. "Making noise about silence: repression of repeated genes in animals." *Current opinion in genetics & development* 10.2 (2000): 211-216.

Bird, Adrian P, and Edwin M Southern. "Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis*." *Journal of molecular biology* 118.1 (1978): 27-47.

Bird, Adrian P, Mary H Taggart, and Barbara A Smith. "Methylated and unmethylated DNA compartments in the sea urchin genome." *Cell* 17.4 (1979): 889-901.

Bird, Adrian P. "CpG-rich islands and the function of DNA methylation." *Nature* 321.6067 (1985): 209-213.

Bird, Adrian P. "CpG islands as gene markers in the vertebrate nucleus." *Trends in Genetics* 3 (1987): 342-347.

Blaschke, Kathryn et al. "Vitamin [thinsp] C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells." *Nature* 500.7461 (2013): 222-226.

Borchert, Glen M., William Lanier, and Beverly L. Davidson. "RNA Polymerase III Transcribes Human MicroRNAs." *Nature Structural & Molecular Biology* 13.12 (2006): 1097-101.

Boroviak, Thorsten et al. "The ability of inner cell mass cells to self-renew as embryonic stem cells is acquired upon epiblast specification." *Nature cell biology* 16.6 (2014): 516.

Bostick, Magnolia et al. "UHRF1 plays a role in maintaining DNA methylation in mammalian cells." *Science* 317.5845 (2007): 1760-1764.

Bourc'his, Déborah et al. "Dnmt3L and the establishment of maternal genomic imprints." *Science* 294.5551 (2001): 2536-2539.

Bourc'his, Déborah, and Timothy H Bestor. "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." *Nature* 431.7004 (2004): 96-99.

Boyes, Joan, and Adrian Bird. "DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein." *Cell* 64.6 (1991): 1123-1134.

Bransteitter, Ronda et al. "Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase." *Proceedings of the National Academy of Sciences* 100.7 (2003): 4102-4107.

Brook, FA, and RL Gardner. "The origin and efficient derivation of embryonic stem cells in the mouse." *Proceedings of the National Academy of Sciences* 94.11 (1997): 5709-5712.

Burdon, Tom et al. "Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells." *Cells Tissues Organs* 165.3-4 (1999): 131-143.

Cadman, Ed, and Chris Benz. "Uridine and Cytidine Metabolism following Inhibition of De Novo Pyrimidine Synthesis by Pyrazofurin." *Biochimica Et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 609.3 (1980): 372-82.

Canham, Maurice A et al. "Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript." *PLoS Biol* 8.5 (2010): e1000379.

Cardoso, M Cristina, and Heinrich Leonhardt. "DNA methyltransferase is actively retained in the cytoplasm during early development." *The Journal of cell biology* 147.1 (1999): 25-32.

Carofino, Brandi L, Bernard Ayanga, and Monica J Justice. "A mouse model for inducible overexpression of Prdm14 results in rapid-onset and highly penetrant T-cell acute lymphoblastic leukemia (T-ALL)." *Disease Models and Mechanisms* 6.6 (2013): 1494-1506.

Carr, Alan, and Mark D Biggin. "A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos." *The EMBO journal* 18.6 (1999): 1598-1608.

Carrozza, Michael J et al. "Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription." *Cell* 123.4 (2005): 581-592.

Cavalier-Smith, Thomas. "Cell volume and the evolution of eukaryotic genome size." *The evolution of genome size* (1985): 105-184.

Cedar, Howard, and Gregory L Verdine. "Gene expression: the amazing demethylase." *Nature* 397.6720 (1999): 568-569.

Chambers, Ian et al. "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." *Cell* 113.5 (2003): 643-655.

Chambers, Ian et al. "Nanog safeguards pluripotency and mediates germline development." *Nature* 450.7173 (2007): 1230-1234.

Chazaud, Claire et al. "Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway." *Developmental cell* 10.5 (2006): 615-624.

Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. "A dicer-independent miRNA biogenesis pathway that requires Ago catalysis." *Nature*. 2010;465:584–9.

Chen, Taiping et al. "Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b." *Molecular and cellular biology* 23.16 (2003): 5594-5605.

Chen, Liguang et al. "Suppression of ES cell differentiation by retinol (vitamin A) via the overexpression of Nanog." *Differentiation* 75.8 (2007): 682-693.

Chen, Liguang, and Jaspal S Khillan. "Promotion of Feeder-Independent Self-Renewal of Embryonic Stem Cells by Retinol (Vitamin A)." *Stem Cells* 26.7 (2008): 1858-1864.

Chen, Lingyi et al. "Cross-regulation of the Nanog and Cdx2 promoters." *Cell research* 19.9 (2009): 1052-1061.

Chen, Liguang, and Jaspal S Khillan. "A Novel Signaling by Vitamin A/Retinol Promotes Self Renewal of Mouse Embryonic Stem Cells by Activating PI3K/Akt Signaling Pathway via Insulin-Like Growth Factor-1 Receptor." *Stem Cells* 28.1 (2010): 57-63.

Chen, Pao-Yang et al. "A comparative analysis of DNA methylation across human embryonic stem cell lines." *Genome biology* 12.7 (2011): 1.

Chen, Tong, Yanan You, Hua Jiang, and Zack Z. Wang. "Epithelial-Mesenchymal Transition (EMT): A Biological Process in the Development, Stem Cell Differentiation and Tumorigenesis." *Journal of Cellular Physiology* (2017).

Cheng, Xiaodong, and Robert M Blumenthal. "Mammalian DNA methyltransferases: a structural perspective." *Structure* 16.3 (2008): 341-350.

Chew, Joon-Lin et al. "Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells." *Molecular and cellular biology* 25.14 (2005): 6031-6046.

Chi, Tian H et al. "Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes." *Nature* 418.6894 (2002): 195-199.

Chung, Tung-Liang et al. "Vitamin C promotes widespread yet specific DNA demethylation of the epigenome in human embryonic stem cells." *Stem Cells* 28.10 (2010): 1848-1855.

Clagett-Dame, Margaret, and Hector F DeLuca. "The role of vitamin A in mammalian reproduction and embryonic development." *Annual review of nutrition* 22.1 (2002): 347-381.

Cokus, Shawn J et al. "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning." *Nature* 452.7184 (2008): 215-219.

Collas, Philippe. "Modulation of plasmid DNA methylation and expression in zebrafish embryos." *Nucleic acids research* 26.19 (1998): 4454-4461.

Conrad, Sabine et al. "Generation of pluripotent stem cells from adult human testis." *Nature* 456.7220 (2008): 344-349.

- Cooper, David Neil. "Eukaryotic DNA methylation." *Human genetics* 64.4 (1983): 315-333.
- Davis, Samuel et al. "LIFR beta and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor." *Science* 260.5115 (1993): 1805-1808.
- Davis, Tamara L et al. "The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development." *Human molecular genetics* 9.19 (2000): 2885-2894.
- Dawlaty, Meelad M et al. "Loss of Tet enzymes compromises proper differentiation of embryonic stem cells." *Developmental cell* 29.1 (2014): 102-111.
- De Napoles, Mariana, Tatyana Nesterova, and Neil Brockdorff. "Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells." *PloS one* 2.9 (2007): e860.
- Deaton, Aimée M et al. "Cell type-specific DNA methylation at intragenic CpG islands in the immune system." *Genome research* 21.7 (2011): 1074-1086.
- del Corral, Ruth Diez et al. "Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension." *Neuron* 40.1 (2003): 65-79.
- Denli, Ahmet M., Bastiaan B. J. Tops, Ronald H. A. Plasterk, René F. Ketting, and Gregory J. Hannon. "Processing of Primary MicroRNAs by the Microprocessor Complex." *Nature* 432.7014 (2004): 231-35.
- Dettman, EJ et al. "Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling common lymphoid progenitors." *Oncogene* 30.25 (2011): 2859-2873.
- Dietrich, Jens-Erik, and Takashi Hiiragi. "Stochastic patterning in the mouse pre-implantation embryo." *Development* 134.23 (2007): 4219-4231.
- Dittrich, Bärbel et al. "Molecular diagnosis of the Prader-Willi and Angelman syndromes by detection of parent-of-origin specific DNA methylation in 15q11-13." *Human genetics* 90.3 (1992): 313-315.
- Do, Dang Vinh et al. "A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo." *Genes & development* 27.12 (2013): 1378-1390.
- Doi, Akiko et al. "Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts." *Nature genetics* 41.12 (2009): 1350-1353.
- Dong, Kevin B et al. "DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity." *The EMBO journal* 27.20 (2008): 2691-2701.
- Duester, Gregg. "Retinoic Acid Synthesis and Signaling during Early Organogenesis." *Cell* 134.6 (2008): 921-31.
- Dunn, S-J et al. "Defining an essential transcription factor program for naive pluripotency." *Science* 344.6188 (2014): 1156-1160.
- Durston, AJ et al. "Retinoic acid causes an anteroposterior transformation in the developing central nervous system." *Development* 103.2 (1989): 140-144.

- Duxbury, K et al. "Naturally occurring isotopes of an analyte can interfere with doubly deuterated internal standard measurement." *Annals of clinical biochemistry* 45.2 (2008): 210-212.
- Eckhardt, Florian et al. "DNA methylation profiling of human chromosomes 6, 20 and 22." *Nature genetics* 38.12 (2006): 1378-1385.
- Esteban, Miguel Angel et al. "Vitamin C enhances the generation of mouse and human induced pluripotent stem cells." *Cell stem cell* 6.1 (2010): 71-79.
- Esteller, Manel. "Epigenetics in cancer." *New England Journal of Medicine* 358.11 (2008): 1148-1159.
- Evans, Martin J, and Matthew H Kaufman. "Establishment in culture of pluripotential cells from mouse embryos." *Nature* 292.5819 (1981): 154-156.
- Evans, D. R., and H. I. Guy. "Mammalian Pyrimidine Biosynthesis: Fresh Insights into an Ancient Pathway." *Journal of Biological Chemistry* 279.32 (2004): 33035-3038.
- Fairbanks, LD. "Importance of Ribonucleotide Availability to Proliferating T-lymphocytes from Healthy Humans." *Journal of Biological Chemistry* 270.50 (1995): 29682-9689.
- Farazi, T. A., S. A. Juranek, and T. Tuschl. "The Growing Catalog of Small RNAs and Their Association with Distinct Argonaute/Piwi Family Members." *Development* 135.7 (2008): 1201-214.
- Farnham, Peggy J. "Insights from genomic profiling of transcription factors." *Nature Reviews Genetics* 10.9 (2009): 605-616.
- Feinberg, Andrew P, and Bert Vogelstein. "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." *Nature* 301.5895 (1983): 89-92.
- Feldmann, Angelika et al. "Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions." *PLoS Genet* 9.12 (2013): e1003994.
- Feng, Jian et al. "Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons." *Nature neuroscience* 13.4 (2010): 423-430.
- Fenn, JB et al. "Electrospray ionization for mass-spectrometry of large biomolecules." *Science* 246.4926 (1989): 64-71.
- Ficz, Gabriella et al. "FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency." *Cell Stem Cell* 13.3 (2013): 351-359.
- Franchini, Don-Marc, and Svend K Petersen-Mahrt. "AID and APOBEC deaminases: balancing DNA damage in epigenetics and immunity." *Epigenomics* 6.4 (2014): 427-443.
- Frankenberg, Stephen et al. "Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling." *Developmental cell* 21.6 (2011): 1005-1013.
- Fujikura, Junji et al. "Differentiation of embryonic stem cells is induced by GATA factors." *Genes & development* 16.7 (2002): 784-789.

- Gama-Sosa, Miguel A et al. "The 5-methylcytosine content of DNA from human tumors." *Nucleic acids research* 11.19 (1983): 6883-6894.
- Gao, Songmei, Zong-Ping Zhang, and HT Karnes. "Sensitivity enhancement in liquid chromatography/atmospheric pressure ionization mass spectrometry using derivatization and mobile phase additives." *Journal of Chromatography B* 825.2 (2005): 98-110.
- Gardner, RL. "Origin and differentiation of extraembryonic tissues in the mouse." *International review of experimental pathology* 24 (1983): 63-133.
- Ginsburg, MALKA, MH Snow, and ANNE McLAREN. "Primordial germ cells in the mouse embryo during gastrulation." *Development* 110.2 (1990): 521-528.
- Gjerset, Ruth A, and DW Martin. "Presence of a DNA demethylating activity in the nucleus of murine erythroleukemic cells." *Journal of Biological Chemistry* 257.15 (1982): 8581-8583.
- Gong, Zhizhong et al. "ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase." *cell* 111.6 (2002): 803-814.
- Gowher, Humaira, Oliver Leismann, and Albert Jeltsch. "DNA of *Drosophila melanogaster* contains 5-methylcytosine." *The EMBO journal* 19.24 (2000): 6918-6923.
- Gowher, Humaira et al. "Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L." *Journal of Biological Chemistry* 280.14 (2005): 13341-13348.
- Gowher, Humaira et al. "Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase." *Journal of molecular biology* 357.3 (2006): 928-941.
- Grabole, Nils et al. "Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation." *EMBO reports* 14.7 (2013): 629-637.
- Grin, Inga, and Alexander A Ishchenko. "An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation." *Nucleic acids research* 44.8 (2016): 3713-3727.
- Grosser, Christian et al. "Altering TET dioxygenase levels within physiological range affects DNA methylation dynamics of HEK293 cells." *Epigenetics* 10.9 (2015): 819-833.
- Gu, Tian-Peng et al. "The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes." *Nature* 477.7366 (2011): 606-610.
- Guibert, Sylvain, Thierry Forné, and Michael Weber. "Global profiling of DNA methylation erasure in mouse primordial germ cells." *Genome research* 22.4 (2012): 633-641.
- Guo, Ge et al. "Klf4 reverts developmentally programmed restriction of ground state pluripotency." *Development* 136.7 (2009): 1063-1069.
- Guo, Guoji et al. "Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to

blastocyst." *Developmental cell* 18.4 (2010): 675-685.

Guo, Junjie U et al. "Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain." *Nature neuroscience* 17.2 (2014): 215-222.

Habibi, Ehsan et al. "Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells." *Cell stem cell* 13.3 (2013): 360-369.

Hackett, Jamie A et al. "Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine." *Science* 339.6118 (2013): 448-452.

Hackett, Jamie A, and M Azim Surani. "DNA methylation dynamics during the mammalian life cycle." *Phil. Trans. R. Soc. B* 368.1609 (2013): 20110328.

Hahn, Matthew W, and Gregory A Wray. "The g-value paradox." *Evolution and Development* 4.2 (2002): 73-75.

Hahnel, Ann C et al. "Two alkaline phosphatase genes are expressed during early development in the mouse embryo." *Development* 110.2 (1990): 555-564.

Hajkova, Petra et al. "Epigenetic reprogramming in mouse primordial germ cells." *Mechanisms of development* 117.1 (2002): 15-23.

Hajkova, Petra et al. "Chromatin dynamics during epigenetic reprogramming in the mouse germ line." *Nature* 452.7189 (2008): 877-881.

Hajkova, Petra et al. "Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway." *Science* 329.5987 (2010): 78-82.

Halket, John M, and Vladimir G Zaikin. "Derivatization in mass spectrometry-7. On-line derivatisation/degradation." *European Journal of Mass Spectrometry* 12.1 (2006): 1-14.

Hanna, Jacob et al. "Direct cell reprogramming is a stochastic process amenable to acceleration." *Nature* 462.7273 (2009): 595-601.

Hanna, Jacob H, Krishanu Saha, and Rudolf Jaenisch. "Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues." *Cell* 143.4 (2010): 508-525.

Hashimoto, Hideharu et al. "The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix." *Nature* 455.7214 (2008): 826-829.

Hata, Kenichiro et al. "Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells." *Molecular reproduction and development* 73.1 (2006): 116-122.

He, Yu-Fei et al. "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." *Science* 333.6047 (2011): 1303-1307.

HEATH, JOHN K, and AUSTIN G SMITH. "Regulatory factors of embryonic stem cells." *J Cell Sci* 1988.Supplement 10 (1988): 257-266.

Hellman, Asaf, and Andrew Chess. "Gene body-specific methylation on the active X chromosome." *science* 315.5815 (2007): 1141-1143.

Henckel, Amandine et al. "Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals." *Human molecular genetics* 18.18 (2009): 3375-3383.

Hendrich, Brian, and Adrian Bird. "Identification and characterization of a family of mammalian methyl-CpG binding proteins." *Molecular and cellular biology* 18.11 (1998): 6538-6547.

Hendrich, Brian et al. "Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development." *Genes & development* 15.6 (2001): 710-723.

Herring, Jason L., Daniel K. Rogstad, and Lawrence C. Sowers. "Enzymatic Methylation of DNA in Cultured Human Cells Studied by Stable Isotope Incorporation and Mass Spectrometry." *Chemical Research in Toxicology* 22.6 (2009): 1060-068.

Hillman, N, MI Sherman, and Chris Graham. "The effect of spatial arrangement on cell determination during mouse development." *Development* 28.2 (1972): 263-278.

Hikasa, Hiroki et al. "Regulation of TCF3 by Wnt-dependent phosphorylation during vertebrate axis specification." *Developmental cell* 19.4 (2010): 521-532.

Hodges, Emily et al. "High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing." *Genome research* 19.9 (2009): 1593-1605.

Holliday, Robin, and John E Pugh. "DNA modification mechanisms and gene activity during development." *Cold Spring Harbor Monograph Archive* 32 (1996): 639-645.

Hon, Gary C et al. "5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation." *Molecular cell* 56.2 (2014): 286-297.

Honjo, Tasuku. "Does AID need another aid?." *Nature immunology* 3.9 (2002): 800-801.

Hoogenraad, N.J., Lee D.C., "Effect of uridine on de novo pyrimidine biosynthesis in rat hepatoma cells in culture" *J Biol Chem.* 1974 May 10;249(9):2763-8.

Hotchkiss, Rollin D. "The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography." *Journal of Biological Chemistry* 175.1 (1948): 315-332.

Howlett, SARAH K, and WOLF Reik. "Methylation levels of maternal and paternal genomes during preimplantation development." *Development* 113.1 (1991): 119-127.

Hsieh, Chih-Lin. "Evidence that protein binding specifies sites of DNA demethylation." *Molecular and cellular biology* 19.1 (1999): 46-56.

Huang, Yun, and Anjana Rao. "Connections between TET proteins and aberrant DNA modification in cancer." *Trends in Genetics* 30.10 (2014): 464-474.

- Huh, Jin Hoe et al. "Cellular programming of plant gene imprinting." *Cell* 132.5 (2008): 735-744.
- Illingworth, Robert S, and Adrian P Bird. "CpG islands—'a rough guide'." *FEBS letters* 583.11 (2009): 1713-1720.
- Illingworth, Robert S et al. "Orphan CpG islands identify numerous conserved promoters in the mammalian genome." *PLoS Genet* 6.9 (2010): e1001134.
- Ingham, PW. "A clonal analysis of the requirement for the trithorax gene in the diversification of segments in *Drosophila*." *Journal of embryology and experimental morphology* 89 (1985): 349-365.
- Inoue, Azusa, and Yi Zhang. "Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos." *Science* 334.6053 (2011): 194-194.
- Iqbal, Khursheed et al. "Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine." *Proceedings of the National Academy of Sciences* 108.9 (2011): 3642-3647.
- Iyer, Vishwanath R et al. "Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF." *Nature* 409.6819 (2001): 533-538.
- Jackson, Melany et al. "Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells." *Molecular and cellular biology* 24.20 (2004): 8862-8871.
- Jackson-Grusby, Laurie et al. "Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation." *Nature genetics* 27.1 (2001): 31-39.
- Jacob, François. "The Leeuwenhoek lecture, 1977: mouse teratocarcinoma and mouse embryo." *Proceedings of the Royal Society of London B: Biological Sciences* 201.1144 (1978): 249-270.
- Jia, Da et al. "Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation." *Nature* 449.7159 (2007): 248-251.
- Jiang, Lan et al. "Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos." *Cell* 153.4 (2013): 773-784.
- Jin, Bilian et al. "DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function." *Human molecular genetics* 17.5 (2008): 690-709.
- Jin, Chunlei et al. "TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells." *Nucleic acids research* (2014)
- Johnson, David W. "A modified Girard derivatizing reagent for universal profiling and trace analysis of aldehydes and ketones by electrospray ionization tandem mass spectrometry." *Rapid Communications in Mass Spectrometry* 21.18 (2007): 2926-2932.
- Jones, A Louise, Carole L Thomas, and Andrew J Maule. "De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus." *The EMBO Journal* 17.21 (1998): 6385-6393.

- Jones, M. E. "Pyrimidine Nucleotide Biosynthesis in Animals: Genes, Enzymes, and Regulation of UMP Biosynthesis." *Annual Review of Biochemistry* 49.1 (1980): 253-79.
- Jones, Peter L et al. "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription." *Nature genetics* 19.2 (1998): 187-191.
- Johansson, Britt M, and Michael V Wiles. "Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development." *Molecular and Cellular Biology* 15.1 (1995): 141-151.
- Jost, Jean-Pierre. "Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine." *Proceedings of the National Academy of Sciences* 90.10 (1993): 4684-4688.
- Jouneau, A., C. Ciaudo, O. Sismeiro, V. Brochard, L. Jouneau, S. Vandormael-Pournin, J.-Y. Coppee, Q. Zhou, E. Heard, C. Antoniewski, and M. Cohen-Tannoudji. "Naive and Primed Murine Pluripotent Stem Cells Have Distinct MiRNA Expression Profiles." *Rna* 18.2 (2011): 253-64.
- Kafer, Georgia Rose et al. "5-Hydroxymethylcytosine marks sites of DNA damage and promotes genome stability." *Cell reports* 14.6 (2016): 1283-1292.
- Kafri, Tal et al. "Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line." *Genes & development* 6.5 (1992): 705-714.
- Kagiwada, Saya et al. "Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice." *The EMBO journal* 32.3 (2013): 340-353.
- Kaji, Keisuke, Jennifer Nichols, and Brian Hendrich. "Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells." *Development* 134.6 (2007): 1123-1132.
- Kanda, Akifumi et al. "Establishment of ES cells from inbred strain mice by dual inhibition (2i)." *Journal of Reproduction and Development* 58.1 (2012): 77-83.
- Kanellopoulou, C. "Dicer-deficient Mouse Embryonic Stem Cells Are Defective in Differentiation and Centromeric Silencing." *Genes & Development* 19.4 (2005): 489-501.
- Kang, Minjung et al. "FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse." *Development* 140.2 (2013): 267-279.
- Kangaspeska, Sara et al. "Transient cyclical methylation of promoter DNA." *Nature* 452.7183 (2008): 112-115.
- Kaplan, Tommy et al. "Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development." *PLoS Genet* 7.2 (2011): e1001290.
- Karwacki-Neisius, Violetta et al. "Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog." *Cell stem cell* 12.5 (2013): 531-545.
- Kawasaki, Hiroshi et al. "Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity." *Neuron* 28.1 (2000): 31-40.

Kebarle, Paul, and Liang Tang. "From ions in solution to ions in the gas phase-the mechanism of electrospray mass spectrometry." *Analytical chemistry* 65.22 (1993): 972A-986A.

Kelly, Kevin F et al. "β-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism." *Cell stem cell* 8.2 (2011): 214-227.

Kemmerich, Kristin et al. "Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil-and UNG-backup uracil-excision activities and increases cancer predisposition of Ung^{-/-} Msh2^{-/-} mice." *Nucleic acids research* 40.13 (2012): 6016-6025.

Keshet, I, J Yisraeli, and H Cedar. "Effect of regional DNA methylation on gene expression." *Proceedings of the National Academy of Sciences* 82.9 (1985): 2560-2564.

Khillan, Jaspal S. "Vitamin A/retinol and maintenance of pluripotency of stem cells." *Nutrients* 6.3 (2014): 1209-1222.

Kienhöfer, Sabine et al. "GADD45a physically and functionally interacts with TET1." *Differentiation* 90.1 (2015): 59-68.

Kleinsmith, Lewis J, Vincent G Allfrey, and Alfred E Mirsky. "Phosphoprotein metabolism in isolated lymphocyte nuclei." *Proceedings of the National Academy of Sciences* 55.5 (1966): 1182-1189.

Koh, Kian Peng et al. "Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells." *Cell stem cell* 8.2 (2011): 200-213.

Koh, Kian Peng, and Anjana Rao. "DNA methylation and methylcytosine oxidation in cell fate decisions." *Current opinion in cell biology* 25.2 (2013): 152-161.

Kolodziejczyk, Aleksandra A et al. "Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation." *Cell stem cell* 17.4 (2015): 471-485.

Kumar, Roshan M., Patrick Cahan, Alex K. Shalek, Rahul Satija, A. Jay Daleykeyser, Hu Li, Jin Zhang, Keith Pardee, David Gennert, John J. Trombetta, Thomas C. Ferrante, Aviv Regev, George Q. Daley, and James J. Collins. "Deconstructing Transcriptional Heterogeneity in Pluripotent Stem Cells." *Nature* 516.7529 (2014): 56-61

Kumar, S., and G. Duester. "Retinoic Acid Controls Body Axis Extension by Directly Repressing Fgf8 Transcription." *Development* 141.15 (2014): 2972-977

Kunath, Tilo et al. "FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment." *Development* 134.16 (2007): 2895-2902.

Kuo, Kenneth C et al. "Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA." *Nucleic acids research* 8.20 (1980): 4763-4776.

Kurimoto, Kazuki et al. "Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice." *Genes & development* 22.12 (2008): 1617-1635.

Lander, Eric S et al. "Initial sequencing and analysis of the human genome." *Nature* 409.6822 (2001): 860-921.

Lane AN, Fan TW- M (2015) Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Research*

43:2466–2485. doi: 10.1093/nar/gkv047

Larger, Patrice J et al. "Ion-suppression effects in liquid chromatography–tandem mass spectrometry due to a formulation agent, a case study in drug discovery bioanalysis." *Journal of pharmaceutical and biomedical analysis* 39.1 (2005): 206-216.

Larsen, Frank et al. "CpG islands as gene markers in the human genome." *Genomics* 13.4 (1992): 1095-1107.

Lawson, Kirstie A et al. "Bmp4 is required for the generation of primordial germ cells in the mouse embryo." *Genes & development* 13.4 (1999): 424-436.

Lyashenko, Natalia. "Role of beta-catenin in self-renewal and differentiation of mouse embryonic stem cells." 2009.

Lyashenko, Natalia et al. "Differential requirement for the dual functions of [beta]-catenin in embryonic stem cell self-renewal and germ layer formation." *Nature cell biology* 13.7 (2011): 753-761.

Laurent, Louise et al. "Dynamic changes in the human methylome during differentiation." *Genome research* 20.3 (2010): 320-331.

Lawson, Kirstie A et al. "Bmp4 is required for the generation of primordial germ cells in the mouse embryo." *Genes & development* 13.4 (1999): 424-436.

Lee, Y. "MicroRNA Maturation: Stepwise Processing and Subcellular Localization." *The EMBO Journal* 21.17 (2002): 4663-670.

Lei, Hong et al. "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells." *Development* 122.10 (1996): 3195-3205.

Leitch, Harry G et al. "Naive pluripotency is associated with global DNA hypomethylation." *Nature structural & molecular biology* 20.3 (2013): 311-316.

Lepesheva, Galina et al. "Sterol 14 α -demethylase as a potential target for antitrypanosomal therapy: enzyme inhibition and parasite cell growth." *Chemistry & biology* 14.11 (2007): 1283-1293.

Lewis, Benjamin P., Christopher B. Burge, and David P. Bartel. "Conserved Seed Pairing, Often Flanked by Adenosines, Indicates That Thousands of Human Genes Are MicroRNA Targets." *Cell* 120.1 (2005): 15-20.

Li, Zheng et al. "Gadd45a promotes DNA demethylation through TDG." *Nucleic acids research* (2015): gkv283.

Li, En, Timothy H Bestor, and Rudolf Jaenisch. "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." *Cell* 69.6 (1992): 915-926.

Licht, Lawrence E, and Leslie A Lowcock. "Genome size and metabolic rate in salamanders." *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 100.1 (1991): 83-92.

Lin, Iping G, and Chih-Lin Hsieh. "Chromosomal DNA demethylation specified by protein binding." *EMBO reports* 2.2 (2001): 108-112.

Lister, Ryan et al. "Human DNA methylomes at base resolution show widespread epigenomic differences." *nature* 462.7271

(2009): 315-322.

Lluis, Frederic et al. "The Wnt/ β -Catenin Signaling Pathway Tips the Balance Between Apoptosis and Reprogramming of Cell Fusion Hybrids." *Stem Cells* 28.11 (2010): 1940-1949.

Lu, Falong et al. "Role of Tet proteins in enhancer activity and telomere elongation." *Genes & development* 28.19 (2014): 2103-2119.

MacGregor, Grant R, Brian P Zambrowicz, and Philippe Soriano. "Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells." *Development* 121.5 (1995): 1487-1496.

Macrae, Ian J., Kaihong Zhou, and Jennifer A. Doudna. "Structural Determinants of RNA Recognition and Cleavage by Dicer." *Nature Structural & Molecular Biology* 14.10 (2007): 934-40.

Maden, Malcolm, and Nigel Holder. "Retinoic acid and development of the central nervous system." *Bioessays* 14.7 (1992): 431-438.

Maden, Malcolm. "The Role of Retinoic Acid in Embryonic and Post-embryonic Development." *Proceedings of the Nutrition Society* 59.01 (2000): 65-73.

Maiti, Atanu, and Alexander C Drohat. "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine potential implications for active demethylation of CpG sites." *Journal of Biological Chemistry* 286.41 (2011): 35334-35338.

Magnúsdóttir, Erna, and M Azim Surani. "How to make a primordial germ cell." *Development* 141.2 (2014): 245-252.

Margot, Jean B et al. "Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure." *Journal of molecular biology* 297.2 (2000): 293-300.

Marks, Hendrik et al. "The transcriptional and epigenomic foundations of ground state pluripotency." *Cell* 149.3 (2012): 590-604.

Marson, Alexander et al. "Wnt signaling promotes reprogramming of somatic cells to pluripotency." *Cell stem cell* 3.2 (2008): 132.

Marson, Alexander, Stuart S. Levine, Megan F. Cole, Garrett M. Frampton, Tobias Brambrink, Sarah Johnstone, Matthew G. Guenther, Wendy K. Johnston, Marius Wernig, Jamie Newman, J. Mauro Calabrese, Lucas M. Dennis, Thomas L. Volkert, Sumeet Gupta, Jennifer Love, Nancy Hannett, Phillip A. Sharp, David P. Bartel, Rudolf Jaenisch, and Richard A. Young. "Connecting MicroRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells." *Cell* 134.3 (2008): 521-33.

Martello, Graziano et al. "Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal." *Cell stem cell* 11.4 (2012): 491-504.

Martello, Graziano, Paul Bertone, and Austin Smith. "Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor." *The EMBO journal* 32.19 (2013): 2561-2574.

- Martin, Gail R. "Teratocarcinomas as a Model System for Review the Study of Embryogenesis and Neoplasia." (1975).
- Martin, Gail R. "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* 78.12 (1981): 7634-7638.
- Martin, Alberto et al. "Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas." *Nature* 415.6873 (2002): 802-806.
- Martinez-Ceballos, Eduardo, Pierre Chambon, and Lorraine J Gudas. "Differences in gene expression between wild type and Hoxa1 knockout embryonic stem cells after retinoic acid treatment or leukemia inhibitory factor (LIF) removal." *Journal of Biological Chemistry* 280.16 (2005): 16484-16498.
- Martinowich, Keri et al. "DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation." *Science* 302.5646 (2003): 890-893.
- Masui, Shinji et al. "Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells." *Nature cell biology* 9.6 (2007): 625-635.
- Matsuda, Takahiko et al. "STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells." *The EMBO journal* 18.15 (1999): 4261-4269.
- Matuszewski, BK, ML Constanzer, and CM Chavez-Eng. "Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS." *Analytical chemistry* 75.13 (2003): 3019-3030.
- Maunakea, Alikea K et al. "Conserved role of intragenic DNA methylation in regulating alternative promoters." *Nature* 466.7303 (2010): 253-257.
- Mayer, Wolfgang et al. "Embryogenesis: demethylation of the zygotic paternal genome." *nature* 403.6769 (2000): 501-502.
- Meijer, Laurent et al. "GSK-3-selective inhibitors derived from Tyrian purple indirubins." *Chemistry & biology* 10.12 (2003): 1255-1266.
- Meissner, Alexander et al. "Genome-scale DNA methylation maps of pluripotent and differentiated cells." *Nature* 454.7205 (2008): 766-770.
- Menke, Douglas B. *Sexual Differentiation of Somatic and Germ Cell Lineages of the Mouse Gonad*. 2002
- Mertineit, Carmen et al. "Sex-specific exons control DNA methyltransferase in mammalian germ cells." *Development* 125.5 (1998): 889-897.
- Métivier, Raphaël et al. "Cyclical DNA methylation of a transcriptionally active promoter." *Nature* 452.7183 (2008): 45-50.
- Miller, Courtney A, and J David Sweatt. "Covalent modification of DNA regulates memory formation." *Neuron* 53.6 (2007): 857-869.
- Mintz, Beatrice, and Karl Illmensee. "Normal genetically mosaic mice produced from malignant teratocarcinoma cells." *Proceedings of the National Academy of Sciences* 72.9 (1975): 3585-3589.

Mitsui, Kaoru et al. "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." *cell* 113.5 (2003): 631-642.

Monk, Marilyn, and Anne McLaren. "X-chromosome activity in foetal germ cells of the mouse." *Development* 63.1 (1981): 75-84.

Montero, LM et al. "The distribution of 5-methylcytosine in the nuclear genome of plants." *Nucleic acids research* 20.12 (1992): 3207-3210.

Morales-Ruiz, Teresa et al. "Demeter and repressor of silencing 1 encode 5-methylcytosine DNA glycosylases." *Proceedings of the National Academy of Sciences* 103.18 (2006): 6853-6858.

Morano, Annalisa et al. "Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene." *Nucleic acids research* (2014): gkt920.

Morgan, Hugh D et al. "Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues IMPLICATIONS FOR EPIGENETIC REPROGRAMMING." *Journal of Biological Chemistry* 279.50 (2004): 52353-52360.

Morgani, Sophie M et al. "Totipotent embryonic stem cells arise in ground-state culture conditions." *Cell reports* 3.6 (2013): 1945-1957.

Morris, Shelli M et al. "Dual roles for the Dab2 adaptor protein in embryonic development and kidney transport." *The EMBO Journal* 21.7 (2002): 1555-1564.

Müller, Udo et al. "TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation." *Nucleic acids research* (2014): gku552.

Münzel, Martin, Daniel Globisch, Tobias Brückl, Mirko Wagner, Veronika Welzmler, Stylianos Michalakis, Markus Müller, Martin Biel, and Thomas Carell. "Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain." *Angewandte Chemie International Edition* 49.31 (2010): 5375-377.

Murchison, E. P., J. F. Partridge, O. H. Tam, S. Cheloufi, and G. J. Hannon. "Characterization of Dicer-deficient Murine Embryonic Stem Cells." *Proceedings of the National Academy of Sciences* 102.34 (2005): 12135-2140.

Murgatroyd, Chris et al. "Dynamic DNA methylation programs persistent adverse effects of early-life stress." *Nature neuroscience* 12.12 (2009): 1559-1566.

Murray, K. "The occurrence of ϵ -N-methyl lysine in histones." *Biochemistry* 3.1 (1964): 10-15.

Murray, James T et al. "Exploitation of KESTREL to identify NDRG family members as physiological substrates for SGK1 and GSK3." *Biochemical Journal* 384.3 (2004): 477-488.

Nabel, Christopher S et al. "AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation." *Nature chemical biology* 8.9 (2012): 751-758.

Nady, Nataliya et al. "ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation." *Elife* 4 (2015): e10150.

Nakaki, Fumio, and Mitinori Saitou. "PRDM14: a unique regulator for pluripotency and epigenetic reprogramming." *Trends in biochemical sciences* 39.6 (2014): 289-298.

Nashun, Buhe et al. "Continuous histone replacement by Hira is essential for normal transcriptional regulation and de novo DNA methylation during mouse oogenesis." *Molecular cell* 60.4 (2015): 611-625.

Neddermann, Petra, and Josef Jiricny. "Efficient removal of uracil from GU mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells." *Proceedings of the National Academy of Sciences* 91.5 (1994): 1642-1646.

Neri, Francesco et al. "Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics." *Cell reports* 10.5 (2015): 674-683.

Neri, Francesco et al. "TET1 is controlled by pluripotency-associated factors in ESCs and downmodulated by PRC2 in differentiated cells and tissues." *Nucleic acids research* (2015): gkv392.

Ng, Huck-Hui et al. "MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex." *Nature genetics* 23.1 (1999): 58-61.

Nichols, Jennifer et al. "Complementary tissue-specific expression of LIF and LIF-receptor mRNAs in early mouse embryogenesis." *Mechanisms of development* 57.2 (1996): 123-131.

Nichols, Jennifer et al. "Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo." *Development* 136.19 (2009): 3215-3222.

Nichols, Jennifer, and Austin Smith. "Naive and primed pluripotent states." *Cell stem cell* 4.6 (2009): 487-492.

Niederreither, Karen, Peter McCaffery, Ursula C. Dräger, Pierre Chambon, and Pascal Dollé. "Restricted Expression and Retinoic Acid-induced Downregulation of the Retinaldehyde Dehydrogenase Type 2 (RALDH-2) Gene during Mouse Development." *Mechanisms of Development* 62.1 (1997): 67-78.

Nishikawa, Noriko et al. "Gene amplification and overexpression of PRDM14 in breast cancers." *Cancer research* 67.20 (2007): 9649-9657.

Niwa, Hitoshi et al. "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3." *Genes & development* 12.13 (1998): 2048-2060.

Niwa, Hitoshi, Jun-ichi Miyazaki, and Austin G Smith. "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." *Nature genetics* 24.4 (2000): 372-376.

Numata, Shusuke et al. "DNA methylation signatures in development and aging of the human prefrontal cortex." *The American Journal of Human Genetics* 90.2 (2012): 260-272.

Ogawa, Kazuya et al. "Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells." *Biochemical and biophysical research communications* 343.1 (2006): 159-166.

Ohinata, Yasuhide et al. "A signaling principle for the specification of the germ cell lineage in mice." *Cell* 137.3 (2009): 571-584.

Okamura, Katsutomo, Joshua W. Hagen, Hong Duan, David M. Tyler, and Eric C. Lai. "The Mirtron Pathway Generates MicroRNA-Class Regulatory RNAs in *Drosophila*." *Cell* 130.1 (2007): 89-100.

Okano, Masaki, Shaoping Xie, and En Li. "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." *Nature genetics* 19.3 (1998): 219-220.

Okano, Masaki et al. "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." *Cell* 99.3 (1999): 247-257.

Okashita, Naoki et al. "PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells." *Development* 141.2 (2014): 269-280.

Ong, Chin-Tong, and Victor G Corces. "CTCF: an architectural protein bridging genome topology and function." *Nature Reviews. Genetics* 15.4 (2014): 234.

Paling, Nicholas RD et al. "Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling." *Journal of Biological Chemistry* 279.46 (2004): 48063-48070.

Palmieri, Susan L et al. "Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation." *Developmental biology* 166.1 (1994): 259-267.

Panning, Barbara, and Rudolf Jaenisch. "DNA hypomethylation can activate Xist expression and silence X-linked genes." *Genes & development* 10.16 (1996): 1991-2002.

Park, Jong-Eun, Inha Heo, Yuan Tian, Dharendra K. Simanshu, Hyeshik Chang, David Jee, Dinshaw J. Patel, and V. Narry Kim. "Dicer Recognizes the 5' End of RNA for Efficient and Accurate Processing." *Nature* 475.7355 (2011): 201-05.

Parrow, VENDELA et al. "Retinoids are endogenous to the porcine blastocyst and secreted by trophoblast cells at functionally-active levels." *International Journal of Developmental Biology* 42.4 (2004): 629-632.

Pastor, William A et al. "Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells." *Nature* 473.7347 (2011): 394-397. Pastor, William A et al. "Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells." *Nature* 473.7347 (2011): 394-397.

Pellicer, Jaume, Michael F Fay, and Ilia J Leitch. "The largest eukaryotic genome of them all?." *Botanical Journal of the Linnean Society* 164.1 (2010): 10-15.

Penterman, Jon et al. "DNA demethylation in the Arabidopsis genome." *Proceedings of the National Academy of Sciences* 104.16 (2007): 6752-6757.

Petersen-Mahrt, Svend K, Reuben S Harris, and Michael S Neuberger. "AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification." *Nature* 418.6893 (2002): 99-104.

Pfaffeneder, Toni et al. "Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA." *Nature chemical biology* 10.7 (2014): 574-581.

Pierce, G Barry. "Teratocarcinoma: model for a developmental concept of cancer." *Current topics in developmental biology* 2

(1967): 223-246.

Poltoratsky, Vladimir P et al. "Recombinogenic phenotype of human activation-induced cytosine deaminase." *The Journal of Immunology* 172.7 (2004): 4308-4313.

Popp, Christian et al. "Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency." *Nature* 463.7284 (2010): 1101-1105.

Potok, Magdalena E et al. "Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern." *Cell* 153.4 (2013): 759-772.

Pritsker, Moshe et al. "Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells." *Proceedings of the National Academy of Sciences* 103.18 (2006): 6946-6951.

Qin, Weihua, Heinrich Leonhardt, and Garwin Pichler. "Regulation of DNA methyltransferase 1 by interactions and modifications." *Nucleus* 2.5 (2011): 392-402.

Rae, Peter MM, and Robert E Steele. "Absence of cytosine methylation at CCGG and GCGC sites in the rDNA coding regions and intervening sequences of *Drosophila* and the rDNA of other higher insects." *Nucleic acids research* 6.9 (1979): 2987-2995.

Rai, Kunal et al. "DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45." *Cell* 135.7 (2008): 1201-1212.

Raiber, Eun-Ang et al. "5-Formylcytosine alters the structure of the DNA double helix." *Nature structural & molecular biology* 22.1 (2015): 44-49.

Rappolee, Daniel A et al. "Expression and function of FGF-4 in peri-implantation development in mouse embryos." *Development* 120.8 (1994): 2259-2269.

Ramchandani, Shyam et al. "DNA methylation is a reversible biological signal." *Proceedings of the National Academy of Sciences* 96.11 (1999): 6107-6112.

Riccio, Antonio et al. "The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability." *Nature genetics* 23.3 (1999): 266-268.

Richardson, Brian E, and Ruth Lehmann. "Mechanisms guiding primordial germ cell migration: strategies from different organisms." *Nature Reviews Molecular Cell Biology* 11.1 (2010): 37-49.

Riggs, Arthur D. "X inactivation, differentiation, and DNA methylation." *Cytogenetic and Genome Research* 14.1 (1975): 9-25.

Rivera-Pérez, Jaime A, and Terry Magnuson. "Primitive streak formation in mice is preceded by localized activation of *Brachyury* and *Wnt3*." *Developmental biology* 288.2 (2005): 363-371.

Rohwedel, Jürgen, Kaomei Guan, and Anna M Wobus. "Induction of cellular differentiation by retinoic acid in vitro." *Cells Tissues Organs* 165.3-4 (1999): 190-202.

Ross, Sharon A et al. "Retinoids in embryonal development." *Physiological reviews* 80.3 (2000): 1021-1054.

- Rossant, J et al. "Expression of a retinoic acid response element-hsplaZ transgene defines specific domains of transcriptional activity during mouse embryogenesis." *Genes & development* 5.8 (1991): 1333-1344.
- Rossant, Janet, Claire Chazaud, and Yojiro Yamanaka. "Lineage allocation and asymmetries in the early mouse embryo." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 358.1436 (2003): 1341-1349.
- Rountree, Michael R, Kurtis E Bachman, and Stephen B Baylin. "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci." *Nature genetics* 25.3 (2000): 269-277.
- Ruby, J, Graham, Calvin H, Jan, and David P. Bartel. "Intronic MicroRNA Precursors That Bypass Drosha Processing." *Nature* 448.7149 (2007): 83-86
- Sailor, Kurt A., Alejandro F. Schinder, and Pierre-Marie Lledo. "Adult Neurogenesis beyond the Niche: Its Potential for Driving Brain Plasticity." *Current Opinion in Neurobiology* 42 (2017): 111-17.
- Schöler, HR et al. "A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor." *The EMBO journal* 8.9 (1989): 2543.
- Seisenberger, Stefanie et al. "The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells." *Molecular cell* 48.6 (2012): 849-862.
- Sharma, Rohit B, Qingde Wang, and Jaspal S Khillan. "Amplification of tumor inducing putative cancer stem cells (CSCs) by vitamin A/retinol from mammary tumors." *Biochemical and biophysical research communications* 436.4 (2013): 625-631.
- Shibutani, Toshihiro et al. "Guanine-5-carboxylcytosine base pairs mimic mismatches during DNA replication." *Scientific reports* 4 (2014): 5220
- Saladi, Srinivas Vinod, and L Ivana. "ATP dependent chromatin remodeling enzymes in embryonic stem cells." *Stem Cell Reviews and Reports* 6.1 (2010): 62-73.
- Sato, Noboru et al. "Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor." *Nature medicine* 10.1 (2004): 55-63.
- Santos, Fátima et al. "Dynamic reprogramming of DNA methylation in the early mouse embryo." *Developmental biology* 241.1 (2002): 172-182.
- Santos, Fátima et al. "Active demethylation in mouse zygotes involves cytosine deamination and base excision repair." *Epigenetics & chromatin* 6.1 (2013)
- Saxonov, Serge, Paul Berg, and Douglas L Brutlag. "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." *Proceedings of the National Academy of Sciences* 103.5 (2006): 1412-1417.
- Schiesser, Stefan et al. "Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing." *Angewandte Chemie International Edition* 51.26 (2012): 6516-6520.
- Shimosato, Daisuke, Makoto Shiki, and Hitoshi Niwa. "Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells." *BMC developmental biology* 7.1 (2007): 1.

Schmidl, Christian et al. "Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity." *Genome research* 19.7 (2009): 1165-1174.

Schomacher, Lars et al. "Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation." *Nature structural & molecular biology* 23.2 (2016): 116-124.

Seki, Yoshiyuki et al. "Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice." *Development* 134.14 (2007): 2627-2638.

Sérandour, Aurélien A et al. "Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers." *Genome research* 21.4 (2011): 555-565.

Sérandour, Aurélien A et al. "Single-CpG resolution mapping of 5-hydroxymethylcytosine by chemical labeling and exonuclease digestion identifies evolutionarily unconserved CpGs as TET targets." *Genome biology* 17.1 (2016):

Sharif, Jafar et al. "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." *Nature* 450.7168 (2007).

Shen, Li et al. "Genome-wide analysis reveals TET-and TDG-dependent 5-methylcytosine oxidation dynamics." *Cell* 153.3 (2013): 692-706.

Sherman, Michael I. "The culture of cells derived from mouse blastocysts." *Cell* 5.4 (1975): 343-349.

Sherman, Michael I, and Richard A Miller. "F9 embryonal carcinoma cells can differentiate into endoderm-like cells." *Developmental biology* 63.1 (1978): 27-34.

Shi, Yigong, and Joan Massagué. "Mechanisms of TGF- β signaling from cell membrane to the nucleus." *Cell* 113.6 (2003): 685-700.

Shi, Yan et al. "Inducing Embryonic Stem Cells to Differentiate into Pancreatic β Cells by a Novel Three-Step Approach with Activin A and All-Trans Retinoic Acid." *Stem cells* 23.5 (2005): 656-662.

Shibutani, Toshihiro et al. "Guanine-5-carboxylcytosine base pairs mimic mismatches during DNA replication." *Scientific reports* 4 (2014): 5220.

Shiio, Yuzuru, and Robert N Eisenman. "Histone sumoylation is associated with transcriptional repression." *Proceedings of the National Academy of Sciences* 100.23 (2003): 13225-13230.

Shiraki, Toshiyuki et al. "Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage." *Proceedings of the National Academy of Sciences* 100.26 (2003): 15776-15781.

Shukla, Sanjeev et al. "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." *Nature* 479.7371 (2011): 74-79.

Sigoillot, F. D., J. A. Berkowski, S. M. Sigoillot, D. H. Kotsis, and H. I. Guy. "Cell Cycle-dependent Regulation of Pyrimidine Biosynthesis." *Journal of Biological Chemistry* 278.5 (2002): 3403-409.

Silva, Jose, and Austin Smith. "Capturing pluripotency." *Cell* 132.4 (2008): 532-536.

Simpson, Victoria J, Thomas E Johnson, and Richard F Hammen. "Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging." *Nucleic acids research* 14.16 (1986): 6711-6719.

Sinning, Allan R. "Role of Vitamin A in the Formation of Congenital Heart Defects." *The Anatomical Record* 253.5 (1998): 147-53

Smiley, Jeffrey A et al. "Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*." *Biochimica et Biophysica Acta (BBA)-General Subjects* 1723.1 (2005): 256-264.

Smith, Austin G, and Martin L Hooper. "Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells." *Developmental biology* 121.1 (1987): 1-9.

Smith, Austin G et al. "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." *Nature* 336.6200 (1988): 688-690.

Song, Liguang et al. "Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry." *Analytical chemistry* 77.2 (2005): 504-510.

Song, Jikui et al. "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation." *Science* 331.6020 (2011): 1036-1040.

Song, Chun-Xiao, Chengqi Yi, and Chuan He. "Mapping recently identified nucleotide variants in the genome and transcriptome." *Nature biotechnology* 30.11 (2012): 1107-1116.

Stadtfeld, Matthias, and Konrad Hochedlinger. "Induced pluripotency: history, mechanisms, and applications." *Genes & development* 24.20 (2010): 2239-2263.

Stadtfeld, Matthias et al. "Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells." *Nature genetics* 44.4 (2012): 398-405.

Stamos, Jennifer L, and William I Weis. "The β -catenin destruction complex." *Cold Spring Harbor perspectives in biology* 5.1 (2013): a007898.

Stancheva, Irina, Carmel Hensey, and Richard R Meehan. "Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in *Xenopus* embryos." *The EMBO journal* 20.8 (2001): 1963-1973.

Stavridis, Marios P et al. "A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification." *Development* 134.16 (2007): 2889-2894.

Stein, R, Aharon Razin, and Howard Cedar. "In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells." *Proceedings of the National Academy of Sciences* 79.11 (1982): 3418-3422.

Stein, Reuven et al. "Clonal inheritance of the pattern of DNA methylation in mouse cells." *Proceedings of the National Academy of Sciences* 79.1 (1982): 61-65.

Stevens, Leroy C. "Origin of testicular teratomas from primordial germ cells in mice." *Journal of the National Cancer Institute* 38.4 (1967): 549-552.

Stevens, Leroy C. "The development of transplantable teratocarcinomas from intratesticular grafts of pre-and postimplantation mouse embryos." *Developmental biology* 21.3 (1970): 364-382.

Stevens, Leroy C. "A new inbred subline of mice (129/terSv) with a high incidence of spontaneous congenital testicular teratomas." *Journal of the National Cancer Institute* 50.1 (1973): 235-242.

Stevens, Leroy C, and Clarence Cook Little. "Spontaneous testicular teratomas in an inbred strain of mice." *Proceedings of the National Academy of Sciences* 40.11 (1954): 1080-1087.

Stewart, Colin L et al. "Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor." *Nature* 359.6390 (1992): 76-79.

Stokvis, Ellen, Hilde Rosing, and Jos H Beijnen. "Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not?." *Rapid Communications in Mass Spectrometry* 19.3 (2005): 401-407.

Struhl, Kevin. "Histone acetylation and transcriptional regulatory mechanisms." *Genes & development* 12.5 (1998): 599-606.

Strumpf, Dan et al. "Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst." *Development* 132.9 (2005): 2093-2102.

Sun, Hong et al. "PTEN modulates cell cycle progression and cell survival by regulating phosphatidylinositol 3, 4, 5,-trisphosphate and Akt/protein kinase B signaling pathway." *Proceedings of the National Academy of Sciences* 96.11 (1999): 6199-6204.

Sun, Zhiyi et al. "A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution." *Molecular cell* 57.4 (2015): 750-761.

Suttle, D. P., B. Y. Bugg, J. K. Winkler, and J. J. Kanalas. "Molecular Cloning and Nucleotide Sequence for the Complete Coding Region of Human UMP Synthase." *Proceedings of the National Academy of Sciences* 85.6 (1988): 1754-1758.

Swisher, Jennifer FA et al. "Analysis of putative RNase sensitivity and protease insensitivity of demethylation activity in extracts from rat myoblasts." *Nucleic acids research* 26.24 (1998): 5573-5580.

Syeda, Farisa et al. "The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1." *Journal of Biological Chemistry* 286.17 (2011): 15344-15351.

Szathmáry, Eörs, Ferenc Jordán, and Csaba Pál. "Can genes explain biological complexity?." *Science* 292.5520 (2001): 1315-1316.

Tachibana, Makoto et al. "G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription." *The EMBO journal* 27.20 (2008): 2681-2690.

Takahashi-Tezuka, Mariko et al. "Gab1 acts as an adapter molecule linking the cytokine receptor gp130 to ERK mitogen-activated protein kinase." *Molecular and cellular biology* 18.7 (1998): 4109-4117.

Takahashi, Kazutoshi et al. "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." *cell* 131.5 (2007): 861-872.

Takeshita, Kohei et al. "Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1)." *Proceedings of the National Academy of Sciences* 108.22 (2011): 9055-9059.

Tam, Patrick PL, and Sheila X Zhou. "The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo." *Developmental biology* 178.1 (1996): 124-132.

Tang, Yong et al. "Jak/Stat3 signaling promotes somatic cell reprogramming by epigenetic regulation." *Stem Cells* 30.12 (2012): 2645-2656.

Tang, Yong, and Xiuchun Tian. "JAK-STAT3 and somatic cell reprogramming." *Jak-Stat* 2.4 (2013): e24935.

Tarkowski, Andrzej K, and Joanna Wróblewska. "Development of blastomeres of mouse eggs isolated at the 4-and 8-cell stage." *Development* 18.1 (1967): 155-180.

Taylor, Shirley M, and Peter A Jones. "Multiple new phenotypes induced in 10T12 and 3T3 cells treated with 5-azacytidine." *Cell* 17.4 (1979): 771-779.

Taylor, Paul J. "Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry." *Clinical biochemistry* 38.4 (2005): 328-334.

Ten Berge, Derk et al. "Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells." *Nature cell biology* 13.9 (2011): 1070-1075.

Thomson, Matt et al. "Pluripotency factors in embryonic stem cells regulate differentiation into germ layers." *Cell* 145.6 (2011): 875-889.

Toyooka, Yayoi et al. "Identification and characterization of subpopulations in undifferentiated ES cell culture." *Development* 135.5 (2008): 909-918.

Tsumura, Akiko et al. "Maintenance of self–renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b." *Genes to Cells* 11.7 (2006): 805-814.

Tu, Shengjiang et al. "Co-repressor CBFA2T2 regulates pluripotency and germline development." *Nature* (2016).

Valentine, James W. "Two genomic paths to the evolution of complexity in bodyplans." *Paleobiology* 26.03 (2000): 513-519.

Vardimon, Lily et al. "Expression of a cloned adenovirus gene is inhibited by in vitro methylation." *Proceedings of the National Academy of Sciences* 79.4 (1982): 1073-1077.

Venter, J Craig et al. "The sequence of the human genome." *science* 291.5507 (2001): 1304-1351.

Vidali, G, EL Gershey, and VG Allfrey. "Chemical Studies of Histone Acetylation THE DISTRIBUTION OF ε-N-ACETYLLYSINE IN CALF THYMUS HISTONES." *Journal of Biological Chemistry* 243.24 (1968): 6361-6366.

von Meyenn, Ferdinand et al. "Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells." *Molecular cell* (2016).

Wade, Paul A et al. "Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation." *Nature genetics* 23.1 (1999): 62-66.

Wang, Wei et al. "Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1." *Proceedings of the National Academy of Sciences* 108.45 (2011): 18283-18288.

Wang, Yangming, Rostislav Medvid, Collin Melton, Rudolf Jaenisch, and Robert Blelloch. "DGCR8 Is Essential for MicroRNA Biogenesis and Silencing of Embryonic Stem Cell Self-renewal." *Nature Genetics* 39.3 (2007): 380-85.

Walsh, Colum P, J Richard Chaillet, and Timothy H Bestor. "Transcription of IAP endogenous retroviruses is constrained by cytosine methylation." *Nature genetics* 20.2 (1998): 116-117.

Wassenegger, Michael et al. "RNA-directed de novo methylation of genomic sequences in plants." *Cell* 76.3 (1994): 567-576.

Watson, James D, and Francis HC Crick. "Molecular structure of nucleic acids." *Nature* 171.4356 (1953): 737-738.

Watterson, D Martin, David B Iverson, and Linda J Van Eldik. "Rapid separation and quantitation of 3', 5'-cyclic nucleotides and 5'-nucleotides in phosphodiesterase reaction mixtures using high-performance liquid chromatography." *Journal of biochemical and biophysical methods* 2.3 (1980): 139-146.

Weber, Michael et al. "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." *Nature genetics* 39.4 (2007): 457-466.

Weber, Alain R et al. "Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism." *Nature communications* 7 (2016).

Weiss, Alona et al. "DNA demethylation in vitro: involvement of RNA." *Cell* 86.5 (1996): 709-718.

Welham, Melanie J et al. "Controlling embryonic stem cell proliferation and pluripotency: the role of PI3K-and GSK-3-dependent signalling." *Biochemical Society Transactions* 39.2 (2011): 674-678.

Wierzbicki, Andrzej T et al. "RNA polymerase V transcription guides ARGONAUTE4 to chromatin." *Nature genetics* 41.5 (2009): 630-634.

Wicklow, Eryn et al. "HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst." *PLoS Genet* 10.10 (2014): e1004618.

Wiles, Michael V, and Britt M Johansson. "Embryonic stem cell development in a chemically defined medium." *Experimental cell research* 247.1 (1999): 241-248.

Williams, R Lindsay, Douglas J Hiltonl, and Nicos A Nicolai. "Myeloid leukemia inhibitory factor maintains the developmental potential of embryonic stem cells." *Nature* 336 (1988): 15.

Williams, Renee T, and Yinsheng Wang. "A density functional theory study on the kinetics and thermodynamics of

- N-glycosidic bond cleavage in 5-substituted 2'-deoxycytidines." *Biochemistry* 51.32 (2012): 6458-6462.
- Wobus, Anna M et al. "Retinoic acid accelerates embryonic stem cell-derived cardiac differentiation and enhances development of ventricular cardiomyocytes." *Journal of molecular and cellular cardiology* 29.6 (1997): 1525-1539.
- Wolf, Stanley F et al. "Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation." *Proceedings of the National Academy of Sciences* 81.9 (1984): 2806-2810.
- Workman, JL, and RE Kingston. "Alteration of nucleosome structure as a mechanism of transcriptional regulation." *Annual review of biochemistry* 67.1 (1998): 545-579.
- Wossidlo, Mark et al. "5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming." *Nature communications* 2 (2011): 241.
- Wray, Jason, Tuzer Kalkan, and Austin G Smith. "The ground state of pluripotency." *Biochemical Society Transactions* 38.4 (2010): 1027-1032.
- Wray, Jason et al. "Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation." *Nature cell biology* 13.7 (2011): 838-845.
- Wrobel, Karl-Heinz, Daniela Bickel, and Richard Kujat. "Immunohistochemical study of seminiferous epithelium in adult bovine testis using monoclonal antibodies against Ki-67 protein and proliferating cell nuclear antigen (PCNA)." *Cell and tissue research* 283.2 (1996): 191-201.
- Wu, L., J. Fan, and J. G. Belasco. "MicroRNAs Direct Rapid Deadenylation of mRNA." *Proceedings of the National Academy of Sciences* 103.11 (2006): 4034-039.
- Wu, Naijun, and Andrew M. Clausen. "Fundamental and Practical Aspects of Ultrahigh Pressure Liquid Chromatography for Fast Separations." *Journal of Separation Science* 30.8 (2007): 1167-182.
- Wu, Susan C, and Yi Zhang. "Active DNA demethylation: many roads lead to Rome." *Nature reviews Molecular cell biology* 11.9 (2010): 607-620.
- Wu, Hao, and Yi Zhang. "Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation." *Genes & development* 25.23 (2011): 2436-2452.
- Wu, Di et al. "Uracil-DNA glycosylase is involved in DNA demethylation and required for embryonic development in the zebrafish embryo." *Journal of Biological Chemistry* 289.22 (2014): 15463-15473.
- Xie, Mingyi, Mingfeng Li, Anna Vilborg, Nara Lee, Mei-Di Shu, Valeria Yartseva, Nenad Šestan, and Joan A. Steitz. "Mammalian 5'-Capped MicroRNA Precursors That Generate a Single MicroRNA." *Cell* 155.7 (2013): 1568-580.
- Xue, Jian-Huang et al. "Uracil-DNA glycosylase UNG promotes Tet-mediated DNA demethylation." *Journal of Biological Chemistry* 291.2 (2016): 731-738.
- Yamagata, Yoshiaki et al. "Rapid turnover of DNA methylation in human cells." *Epigenetics* 7.2 (2012): 141-145.
- Yamaji, Masashi et al. "PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways

in mouse embryonic stem cells." *Cell stem cell* 12.3 (2013): 368-382.

Yan, Jie, Juleen R Zierath, and Romain Barrès. "Evidence for non-CpG methylation in mammals." *Experimental cell research* 317.18 (2011): 2555-2561.

Yang, Xiaojing et al. "Gene body methylation can alter gene expression and is a therapeutic target in cancer." *Cancer cell* 26.4 (2014): 577-590.

Ye, Shoudong et al. "Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1." *The EMBO journal* 32.19 (2013): 2548-2560.

Yeo, Jia-Chi et al. "Klf2 is an essential factor that sustains ground state pluripotency." *Cell Stem Cell* 14.6 (2014): 864-872.

Yi, Fei et al. "Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal." *Nature cell biology* 13.7 (2011): 762-770.

Yi, R. "Exportin-5 Mediates the Nuclear Export of Pre-microRNAs and Short Hairpin RNAs." *Genes & Development* 17.24 (2003): 3011-016.

Ying, Qi-Long et al. "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3." *Cell* 115.3 (2003): 281-292.

Ying, Qi-Long et al. "The ground state of embryonic stem cell self-renewal." *Nature* 453.7194 (2008): 519-523.

Yoder, Jeffrey A, and Timothy H Bestor. "A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast." *Human molecular genetics* 7.2 (1998): 279-284.

Yoshida, Kanji et al. "Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways." *Mechanisms of development* 45.2 (1994): 163-171.

Young, Juan I, Stephan Züchner, and Gaofeng Wang. "Regulation of the Epigenome by Vitamin C." *Annual review of nutrition* 35 (2015): 545.

Yuan, Huabing et al. "Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3." *Genes & development* 9.21 (1995): 2635-2645.

Zhu, Jianhua et al. "The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis." *Current Biology* 17.1 (2007): 54-59.

Zhu, Jiang et al. "On the nature of human housekeeping genes." *Trends in genetics* 24.10 (2008): 481-484.

Zilberman, Daniel, Xiaofeng Cao, and Steven E Jacobsen. "ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation." *Science* 299.5607 (2003): 716-719.

Ziller, Michael J et al. "Genomic distribution and inter-sample variation of non-CpG methylation across human cell types." *PLoS Genet* 7.12 (2011): e1002389.

Zhang, H, and J-K Zhu. "Active DNA demethylation in plants and animals." *Cold Spring Harbor symposia on quantitative*

biology 1 Jan. 2012: 161-173.

Zhang, Tiehong et al. "High expression of PRDM14 correlates with cell differentiation and is a novel prognostic marker in resected non-small cell lung cancer." *Medical Oncology* 30.3 (2013): 1-7.

Zhao, Dayao, Peter Mccaffery, Kathryn J. Ivins, Rachael L. Neve, Patrick Hogan, William W. Chin, and Ursula C. Dräger. "Molecular Identification of a Major Retinoic-Acid-Synthesizing Enzyme, a Retinaldehyde-Specific Dehydrogenase." *European Journal of Biochemistry* 240.1 (1996): 15-22

Zhen, Y et al. "Indirubin-3'-monoxime inhibits autophosphorylation of FGFR1 and stimulates ERK1/2 activity via p38 MAPK." *Oncogene* 26.44 (2007): 6372-6385.

Ziomek, CA, and MH Johnson. "Cell surface interaction induces polarization of mouse 8-cell blastomeres at compaction." *Cell* 21.3 (1980): 935-942.

Zinzen, Robert P et al. "Combinatorial binding predicts spatio-temporal cis-regulatory activity." *Nature* 462.7269 (2009): 65-70.

Zwaka, Thomas P, and James A Thomson. "A germ cell origin of embryonic stem cells?." *Development* 132.2 (2005): 227-233.