Investigation into the molecular triggers of epigenetic reprogramming and meiotic entry in mouse primordial germ cells

Peter William Sloan Hill

Imperial College London
Institute of Clinical Sciences

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This thesis is the result of my own work except where specifically referenced in the text. The work in this thesis was completed under the supervision of Dr Petra Hajkova.

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.

Acknowledgments

The work I present in this thesis wouldn't have been possible without the help and guidance of countless people. To all of them, I say thank you. I've had the great benefit of working alongside many talented scientists, and have learnt something from all of them over the past four years.

First and foremost, I need to acknowledge Petra. To this day, I'm not entirely sure what she saw in the arrogant twenty year-old she first met, but I'm glad she saw something. I wouldn't have become the scientist I am today if she hadn't pushed me. She made me think critically about the most puzzling observations, and together I believe we have constructed models that will hopefully stand the test of time. I couldn't have asked for more support or better guidance, in both life and in science. I hope that many years from now we will still be discussing both over a glass of the finest whiskey.

I would also like to thank all the current and past members of the Reprogramming and Chromatin Group; Petra seems to have an eye for people who are invariably both good scientists and good people. Kirsten, Harry, Zelpha, Sarah and Ola have always been exceedingly kind and always willing to discuss ideas, and have also always known when to put me in my place when need be. Poor Theresa has had to share both a bench and a desk with me over the past few years, and I'm not exactly sure how she has coped with the sprawling mess that has more than occasionally spread onto her side. I've always been glad to be able to discuss baseball and basketball with Tien-Chi, each time reminding me a little bit of my North American roots. In particular, I would like to thank Buhe and Rachel. I've collaborated with both extensively over the past four years, and it has been a real pleasure. And I've also always appreciated being able to practice my rusty French with Rachel as we gossiped about life.

In no particular order, I would also like to thank: Megan, Justyna, Andy and Darren for taking care of my mice when I didn't have the time; James, Magda and Thomas for spending countless hours helping me collect cells; Gopu and Yi for their patience while I frequently pestered them about the various problems I had with the computational work; and the mice, without whom this work would never have been possible. I also should thank Sriharsa, Till, Ana, and Richard who have acted as mentors in some way or another over the past few years.

Finally, I'd like to say a big thank you to Mum, Dad, and Georgia. I know that I haven't been able to spend nearly as much time with all of you as I would have liked while I've been finishing this thesis – and I certainly haven't been back to Canada nearly as much as we all would have liked me to, Mum and Dad. But I do know just how supportive you have all been, so thank you very much for putting up with me!

Abstract

Germ cells represent the link between generations. At the time of fertilisation, sperm and egg come together to form a single cell, called the zygote, from which an entire organism is ultimately constructed. In mouse, two unique properties underlie the exceptional potential of gametes: 1) the ability to execute meiosis; and 2) the correct establishment of complementary epigenetic signatures (genomic imprints) in sperm and oocyte that enable post-implantation embryonic development in the next generation. At the molecular level, both the progression through meiosis and the establishment of sex-specific methylation imprints are critically dependent on a wave of global DNA demethylation that is initiated in primordial germ cells (PGCs), the embryonic precursors of the gametes, following their colonisation of the embryonic gonads. The work I present in this thesis focuses on the molecular mechanisms regulating this wave of DNA demethylation. Specifically, I conclusively show that DNA demethylation proceeds independent of the 5-methylcytosine oxygenase Tet1, previously proposed to act as the putative DNA demethylase. Rather, I show that, with respect to DNA demethylation, Tet1 appears to protect the hypo-methylated state in PGCs following the major loss of 5-methylcytosine. Most strikingly, however, with respect to germ cell development, Tet1 appears to play an important role in regulating the timing of the major developmental and transcriptional transitions that occurs in the mouse foetal germ line between embryonic day (E) 10.5 and E14.5, with PGCs derived from *Tet1*-knockout embryos showing increased stability of the demethylated pre-meiotic germ cell state. Beyond understanding the role of Tet1 in gonadal PGCs, I also present results focusing on the potential relationship between cell signalling, DNA demethylation and licensing for gametogenesis, with preliminary data suggesting a potential role for Akt-mTORC1 signalling in activating the expression of the key meiotic regulator Dazl.

Table of contents

Declaration 1
Acknowledgments 2
Abstract 3
Table of contents 4
Frequently used abbreviations 7

Chapter 1: General Introduction - On the role of gonadal DNA demethylation during the development of the mouse foetal germ line 9

1.1 The unique properties of the mammalian gamete 10
1.1.1 Executing meiosis 10
1.1.2 Sex-specific methylation at imprint control regions enables post-implantation embryonic development 13
1.2 Overview of the germ cell cycle in mouse 17
1.2.1 The murine germ line is specified by induction from pluripotent epiblast cells 17
1.2.2 Transcriptional changes upon PGC specification 19
1.2.3 Migration of PGCs into the genital ridge follows specification 22
1.2.4 Initiation of sex-specific gonocyte differentiation programmes in mouse PGCs 23
1.3 Epigenetics and the acquisition of the unique properties of murine gametes 26
1.3.1 Overview of epigenetic reprogramming in mouse primordial germ cells 26
1.3.2 The importance of epigenetic reprogramming for germ cell development 28
1.4 Developmentally-associated global DNA demethylation events in mammals 33
1.4.1 Potential mechanisms of DNA demethylation 34
1.4.2 The DNA demethylation/re-methylation cycle in mammals 38
1.5 Research focus: in search of the molecular triggers and mechanisms driving DNA demethylation during gonadal epigenetic reprogramming in mouse PGCs 46

Chapter 2: Material and Methods 47

2.1 Mice 48
2.2 Immunostaining 49
2.2.1 Immunostaining of collected cryosections 49
2.2.2 Immunostaining of dissociated cells 51
2.3 PGC isolation by flow cytometry 52
2.4 Next generation sequencing library preparation 53

2.4.1 RNA-Seq library prerparation 53
2.4.2 AbaSeq library preparation 53
2.4.3 Whole genome bisulphite sequencing (WGBS) library preparation 54
2.4.4 Reduced representation bisulphite sequencing (RRBS) library preparation 55
2.5 Bioinformatics 56
2.5.1 Published datasets used 56
2.5.2 RNA-Seq alignment and downstream analysis 56
2.5.3 Whole genome bisulphite sequencing (WGBS) and Tet-assisted bisulphite sequencing (TAB-Seq) alignment and downstream analysis 57
2.5.4 AbaSeq alignment and downstream analysis 61
2.5.5 Reduced representation bisulphite sequencing (RRBS) alignment and downstream analysis 64
2.5.6 ChIP-Seq, hMeDIP-Seq and MeDIP-Seq alignment and downstream analysis 65
2.5.7 Methylation-assisted bisulphite sequencing (MAB-Seq) downstream analysis 67
2.6 Mass spectrometry 67
2.7 Organ culture 68

Chapter 3: Results - On the role of Tet1 during gonadal DNA demethylation in mouse primordial germ cells 69

3.1 Introduction 70
3.2 Results 71
3.2.1 Re-evaluating 5hmC dynamics during gonadal DNA demethylation using integrative single nucleotide resolution DNA modification mapping and quantitative mass spectrometry 71
3.2.2 Tet1 is not required for global DNA demethylation in gonadal PGCs, but is subsequently needed at a small subset of loci for removal of residual methylation and/or aberrant *de novo* methylation 91
3.3 Discussion 100
3.3.1 Gonadal DNA demethylation in PGCs proceeds independent of Tet1 100
3.3.2 5hmC mapping in wild type PGCs and methylation dynamics in Tet1-KO PGCs suggest Tet1 acts predominantly to maintain a hypo-methylated state in PGCs 102

Chapter 4: Results - On the mechanism of Tet1 transcriptional regulation in the mouse foetal germ line 104

4.1 Introduction 105
4.1.1 Tet1 regulates expression of meiosis genes in mouse PGCs 105
4.1.2 Potential mechanisms of Tet1 transcriptional regulation 107
4.2 Results 111

4.2.1 Loss of Tet1 results in increased stability of the demethylated pre-meiotic germ cell state 111

4.2.2 Identification and characterisation of conserved de-regulated Tet1-target genes associated with increased stability of the demethylated pre-meiotic germ cell state 123

4.3 Discussion 132

4.3.1 Re-evaluating the role of Tet1 during transcriptional transitions in mouse foetal germ line 132

4.3.2 Tet1 regulates transcription via both enzymatic and non-enzymatic activity 134

Chapter 5: Results - On the role of Akt-mTORC1 signalling during epigenetic reprogramming and licensing for gametogenesis in the mouse foetal germ line…….…….. 137

5.1 Introduction 138

5.1.1 Renewed investigation into the molecular triggers of gonadal DNA demethylation in mouse PGCs 138

5.1.2 The central role of signalling in mouse PGC development 138

5.1.3 Current understanding for the role of signalling during gonadal epigenetic reprogramming in mouse PGCs 139

5.2 Results 142

5.2.1 RNA-Seq profiling between E8.5 and E12.5 reveals transcriptional changes concurrent with the initiation of gonadal DNA demethylation in mouse PGCs 142

5.2.2 A signalling screen identifies activation of the Akt-mTORC1 signalling axis in PGCs during gonadal epigenetic reprogramming and licensing for gametogenesis 147

5.2.3 Inhibition of mTORC1 signalling does not impair global DNA demethylation but causes significantly reduced expression of Dazl protein 154

5.3 Discussion 157

Chapter 6: .....General Discussion ……..159

6.1 On the mechanism of gonadal DNA demethylation in mouse primordial germ cells 160

6.2 On the timing and molecular triggers of gonadal DNA demethylation and licensing for gametogenesis in mouse PGCs 164

Bibliography ……..168

Frequently used abbreviations

5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
BSA	bovine serum albumin
ChIP-Seq	chromatin immunoprecipitation and sequencing
cUMR	comparatively unmethylated region
DAPI	4',6-diamidino-2-phenylindole
DMEM/F-12	Dulbecco's modified eagle medium: nutrient mixture F-12
E10.5, *etc.*	embryonic day 10.5, *etc.*
EDTA	ethylene diamine tetraacetic acid
EGC	embryonic germ cell
EGTA	ethylene glycol tetraacetic acid
ESC	embryonic stem cell
FBS	foetal bovine serum
FPKM	fragments per kilobase per million
GOF18ΔPE-EGFP	transgenic mouse line with an enhanced green fluorescent protein transgene driven by the *Pou5f1* distal enhancer
H2AK119ub1	histone 2A lysine 119 mono-ubiquitination
H3K4me3	histone 3 lysine 4 tri-methylation
H3K9Ac	histone 3 lysine 9 acetylation
H3K9me2	histone 3 lysine 9 di-methylation
H3K9me3	histone 3 lysine 9 tri-methylation
H3K27Ac	histone 3 lysine 27 acetylation
H3K27me3	histone 3 lysine 27 tri-methylation
H3S10P	histone 3 serine 10 phosphorylation
H4R3me2s	histone 4 arginine 3 symmetrical di-methylation
HCP	high-CpG promoter
hMeDIP-Seq	hydroxymethylated DNA immunoprecipitation and sequencing
ICP	intermediate-CpG promoter
ICR	imprint control region
LC/MS	liquid chromatography mass spectrometry
LCP	low-CpG promoter
MeDIP-Seq	methylated DNA immunoprecipitation and sequencing

PC1, *etc.*	principle component 1, *etc.*
PGC	primordial germ cell
PGCLC	PGC-like cell
PBS	phosphate buffered saline
PFA	paraformaldehyde
RNA-Seq	messenger RNA sequencing
RRBS	reduced representation bisulphite sequencing
TAB-Seq	Tet-assisted bisulphite sequencing
Tet1*-KO, *etc.	*Tet1*-knockout, *etc.*
WGBS	whole genome bisulphite sequencing

Chapter 1

General Introduction

On the role of gonadal

DNA demethylation during the development of

the mouse foetal germ line

1.1 The unique properties of the mammalian gamete

Germ cells are highly specialised bipotent cells, best known as sperm (in males) and oocyte (in females). In mammals, when sperm fertilises oocyte, the result is a one-cell embryo, called a zygote, capable of initiating embryonic and extra-embryonic differentiation programs, and thus ultimately enabling construction of an entire organism (Figure 1.3). In contrast to the germ line, the role of all other embryonic lineages, the soma, is to ensure the appropriate establishment, maintenance, and protection of germ cells so that the genetic information can be passed on from generation to generation. As such, the germ line is unique among cell lineages in its ability to create life, and is thus at the heart of reproduction in multicellular sexually reproducing organisms.

1.1.1 Executing meiosis

In mouse, two properties, among others, underlie the exceptional potential of gametes. First, germ cells execute meiosis, a process during which the initially diploid germ cells transition to a haploid state (Figure 1.1). Thus, by the completion of meiosis, sperm and oocyte contain functionally half the DNA normally contained within a somatic cell. Meiosis itself is preceded by DNA replication, and is initiated with an extended prophase, during which homologous chromosomes pair and undergo recombination in a stepwise manner (Handel and Schimenti, 2010). Two divisions follow the completion of meiotic prophase. In the first division, the homologous chromosomes are separated, generating a cell with two sister chromatids; in the second, the sister chromatids are divided, ultimately generating a haploid gametocyte (Figure 1.1) (Handel and Schimenti, 2010).

Progression through meiosis is a highly conserved feature of germ cells, seen in all eukaryotic organisms undergoing sexual reproduction, and consistently, it is critically important for sexual reproduction. From an evolutionary perspective, the importance of meiosis was highlighted in an

elegant study by Paland and Lynch in which the authors compared the ratio of the number of nucleotide substitutions resulting in an amino acid substitution (K_A) over the number of silent nucleotide substitutions (K_S) for the sexually reproducing (i.e. involving meiosis) and asexually reproducing (i.e. not involving meiosis) forms of the water flea *Daphnia pulex* (Paland and Lynch, 2006). Their results showed that the K_A/K_S ratio was much higher for the asexual branches, where nearly 18% of amino-acid substitutions persist despite being deleterious (compared with under 5% for sexual lineages) (Paland and Lynch, 2006), thus providing strong evidence that the elimination of mildly deleterious mutations from a population is a key advantage of sex, and more specifically, of meiosis (Paland and Lynch, 2006).

Beyond its evolutionary implications, the correct execution of meiosis in gametes is also critical for the subsequent normal embryonic development of the newly formed zygote following fertilisation (Dokshin et al., 2013). For example, in the absence of correct meiotic recombination, random disjunction and aneuploidy can lead to embryonic death or developmental defects (Figure 1) (Hassold et al., 2007).

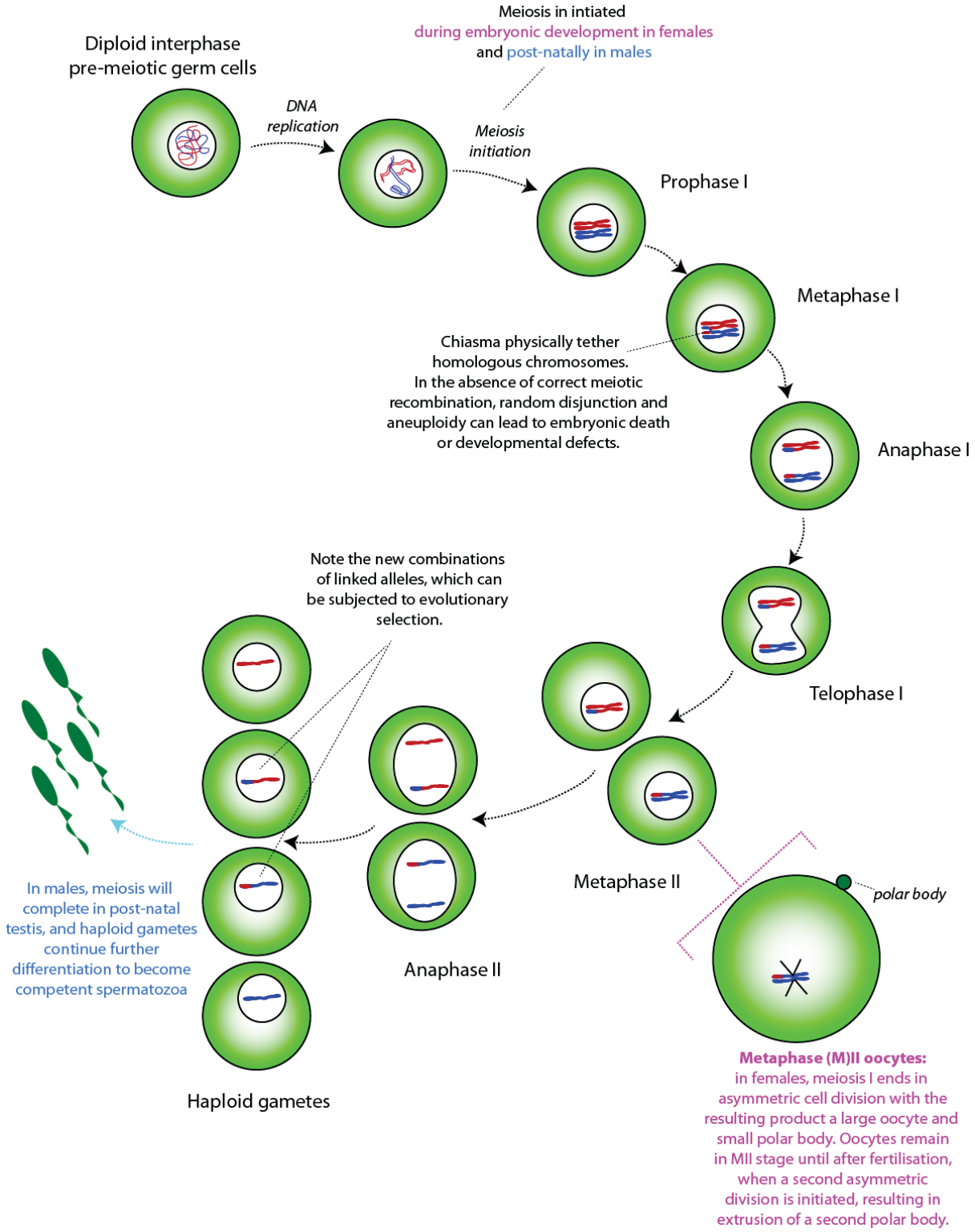


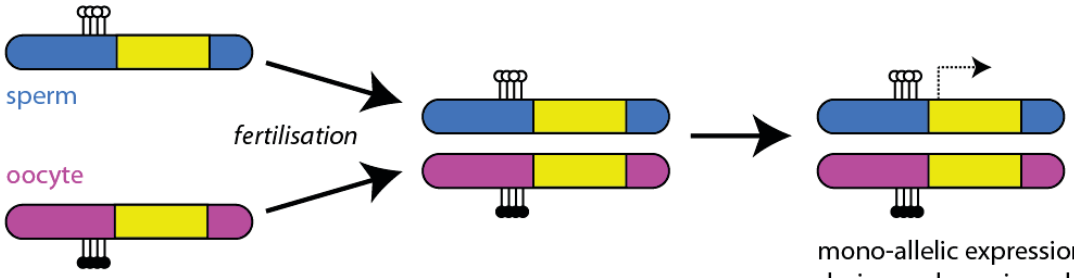
Figure 1.1 – Overview of key meiotic stages in mouse - adapted from (Handel and Schimenti, 2010).

1.1.2 Sex-specific methylation at imprint control regions enables post-implantation embryonic development

The second unique property of mammalian gametes is the correct establishment of complementary epigenetic signatures (genomic imprints) in sperm and oocyte that enable post-implantation embryonic development in the next generation (McGrath and Solter, 1984; Surani et al., 1984). Broadly speaking, epigenetics refers to any mechanisms or phenomena that define a given cell's heritable phenotype in the absence of changes to its genotype (Goldberg et al., 2007). At the molecular level, this includes changes to DNA modifications, histone modifications and variants, nucleosome positioning, and higher order chromatin structure. The best understood aspect of the epigenetic signature observed in mammalian gametes is the sex specific deposition of DNA methylation at imprint control regions (Figure 1.2).

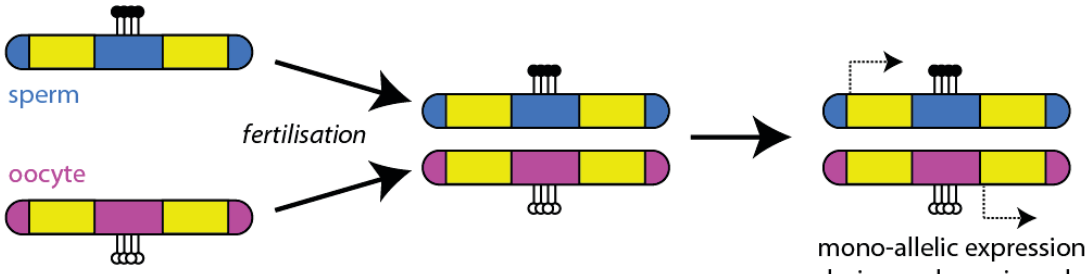
Genomic imprinting refers to the phenomenon whereby genes are expressed in a parent-of-origin specific manner (Tilghman, 1999). Conceptually, the modern understanding of genomic imprinting was first suggested based on a series of elegant experiments by Surani and colleagues and Solter and colleagues. In these experiments, the authors showed that fertilized mouse oocytes manipulated either by removing and replacing the paternal pronucleus with a second maternal one to generate a genetically bi-maternal (i.e. gynogenetic, parthenogenetic) zygote, or, alternatively, by removing and replacing the maternal pronucleus with a second paternal one to generate a genetically bi-paternal (i.e. androgenetic) zygote, could not develop past mid-gestation, with embryos dying shortly after implantation (Figure 1.2C) (McGrath and Solter, 1984; Surani et al., 1984).

A Maternally methylated ICRs are located at imprinted gene promoters



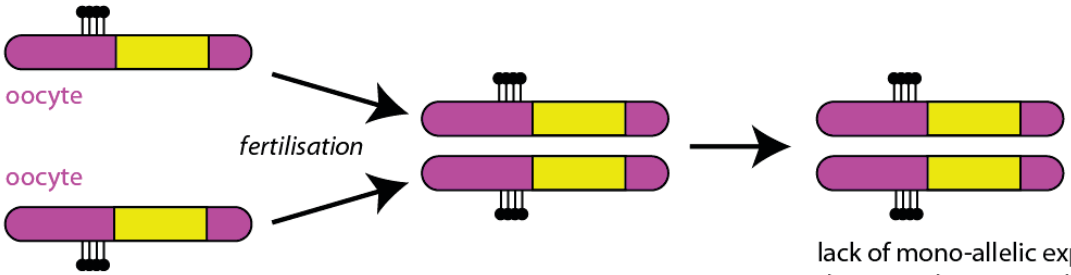
mono-allelic expression
during embryonic and
extra-embryonic development

B Paternally methylated ICRs are located in intergenic regions



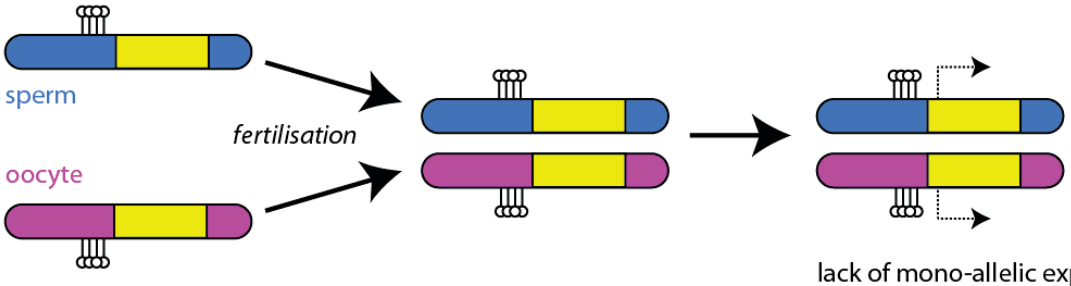
mono-allelic expression
during embryonic and
extra-embryonic development

C Lack of methylation imprinting in parthenogenesis results in embryonic lethality



lack of mono-allelic expression
during embryonic and
extra-embryonic development
results in embryonic lethality

D Lack of *de novo* methylation at ICRs during gametogenesis results in embryonic lethality



lack of mono-allelic expression
during embryonic and
extra-embryonic development
results in embryonic lethality

Figure 1.2 – Overview of methylation imprinting in mouse - adapted from (Bartolomei and Ferguson-Smith, 2011). A) *The majority of imprinted loci are associated with maternal-specific methylation, where DNA methylation (black lollipops) is deposited at the gene promoter of the repressed maternal allele but not the hypo-methylated (white lollipops) expressed (arrow) paternal allele.* ***B)*** *A minority of imprinted loci are methylated on the paternal allele. These tend to be intergenic regions, which regulate a number of imprinted genes. Unlike maternally methylated ICRs, methylation at paternal ICRs can result in both repression and expression of associated imprinted genes.* ***C-D)*** *Lack of methylation imprinting in gametes results in embryonic lethality in the next generation due to aberrant dosage of imprinted gene expression in both embryonic and extra-embryonic lineages.*

The first descriptions of the phenomenon of genomic imprinting set off a search to understand its regulatory mechanisms. Early on it was recognised that imprinting is a consequence of specific genomic loci behaving differently depending on their parental origin, rather than the entire genome behaving in this way, as murine zygotes with maternal or paternal disomy for chromosomes 1, 4, 5, 9, 13, 14 and 15 survive normally, while maternal disomy 6 was shown to be lethal and regions of chromosome 2, 8, 11, 13 and 17, among others, showing non-complementation of maternal duplication/paternal deficiency (Cattanach and Kirk, 1985). While these observations provided useful insight, further mechanistic understanding was only possible following the first characterisation of imprinted genes (Barlow et al., 1991; DeChiara et al., 1991; Ferguson-Smith et al., 1991; Reik et al., 1987), enabling a targeted investigation of the epigenetic state of associated regulatory sequences. Remarkably these studies revealed a relationship between sex-specific CpG methylation and chromatin accessibility/gene expression, hinting at a regulatory role for DNA methylation (Figure 1.2A) (Bartolomei, 1993; Ferguson-Smith et al., 1993; Stöger et al., 1993). This regulatory role was functionally validated by Jaenisch and colleagues, who observed de-regulated expression of all tested (i.e. three) imprinted genes in *Dnmt1*-knockout embryos (Li et al., 1993). Of note, the authors observed activation of the normally silent paternal allele of the *H19* gene, whereas the normally active paternal allele of the *Igf2* gene and maternal allele of the *Igf2r* gene were repressed in mutant embryos (Li et al., 1993), hinting at the complex relationship between DNA methylation and imprinted gene expression (Figure 1.2B).

Subsequent transgenic studies ultimately confirmed that expression of imprinted genes represents the sole barrier to parthenogenetic embryo development (Kono et al., 2004), and that DNA methylation patterns acquired during gametogenesis are critical for the establishment of genomic imprinting (Bourc'his, 2004; Bourc'his et al., 2001; Kaneda et al., 2004). The former was shown by elegant work from Kono and colleagues, in which the authors observed live birth of mice developed from reconstructed oocytes containing two haploid sets of the maternal genome, one of a wild type fully grown oocyte (i.e. containing female-specific methylation imprints) and one from a non-growing oocyte (i.e. lacking all methylation imprints) genetically engineered to enforce the normal expression of only two imprinted genes, *H19* and *Igf2* (Kono et al., 2004). While these remarkable results confirmed that lack of paternal methylation imprinting represents the sole barrier to parthenogenesis (Kono et al., 2004), it should be noted that much more extensive genetic engineering with respect to imprinted gene expression would likely be required for the successful post-implantation development of embryos completely lacking both paternal *and* maternal methylation imprinting.

The observation that DNA methylation patterns acquired during gametogenesis are critical for imprinted gene expression in the next generation was further demonstrated by a series of experiments using transgenic mice depleted of the *de novo* DNA methyltransferase machinery. These studies showed that offspring of *Dnmt3L* knockout females or of females with oocyte-specific deletion of *Dnmt3a* die before mid-gestation as a result of bi-allelic expression of imprinted genes normally methylated on the allele of maternal origin (Figure 1.2D) (Bourc'his et al., 2001; Kaneda et al., 2004). Similarly, loss of Dnmt3a and Dnmt3l results in aberrant methylation imprinting at the paternally methylated imprint control regions in male germ cells (Bourc'his, 2004; Kaneda et al., 2004). In contrast to females, however, genetic knockout of these proteins results in severe meiotic defects during sperm development, and thus no embryonic development in the next generation can be observed at all (Bourc'his, 2004; Kaneda et al., 2004). It should also be noted that one of the three known paternally methylated imprint control regions (*Rasgrf1*) has been shown to be methylated by Dnmt3b in addition to Dnmt3a, along with their co-factor, Dnmt3L (Kato et al., 2007a).

1.2 Overview of the germ cell cycle in mouse

1.2.1 The murine germ line is specified by induction from pluripotent epiblast cells

In sexually reproducing organisms, the germ cell cycle provides the link between generations (Figure 1.3). Following the creation of a new embryo at the time of fertilisation, the process of forming germ cells and soma is started anew, thus enabling the next generation to reproduce itself. Depending on the species, germ line specification is traditionally classified as occurring by either preformation or induction (Lesch and Page, 2012). Preformation necessitates the inheritance of germ cell determinants from oocyte, and thus germ cells and soma are continuously segregated both from the very start of embryogenesis within a given individual and across generations (Lesch and Page, 2012). In contrast, induction involves the existence of pluripotent precursor cells, capable of generating both germ cells and soma, from which the germ line is subsequently segregated during development (Lesch and Page, 2012).

In mouse, all cells of the embryo are capable of contributing to the germ line until after the blastocyst stage. This was shown by elegant studies demonstrating that labelled 4- or 8- cell blastomeres can give rise to both PGCs and somatic cells (Kelly, 1977), and that single pre-implantation epiblast cells injected into donor blastocysts can produce clones contributing to both germ cells and somatic cells, but never just to the germ line (Gardner et al., 1985). The lack of maternally inherited cytoplasmic factors pre-determining germ cell fate was confirmed in a seminal study by Lawson and Hage, who showed that single proximal epiblast cells at embryonic day (E) 6.25 produced eight clones that contributed to the germ cell lineage (Lawson and Hage, 1994); importantly, however, all of these clones also contributed to somatic lineages, and thus no germ lineage-restricted epiblast cells were found (Lawson and Hage, 1994).

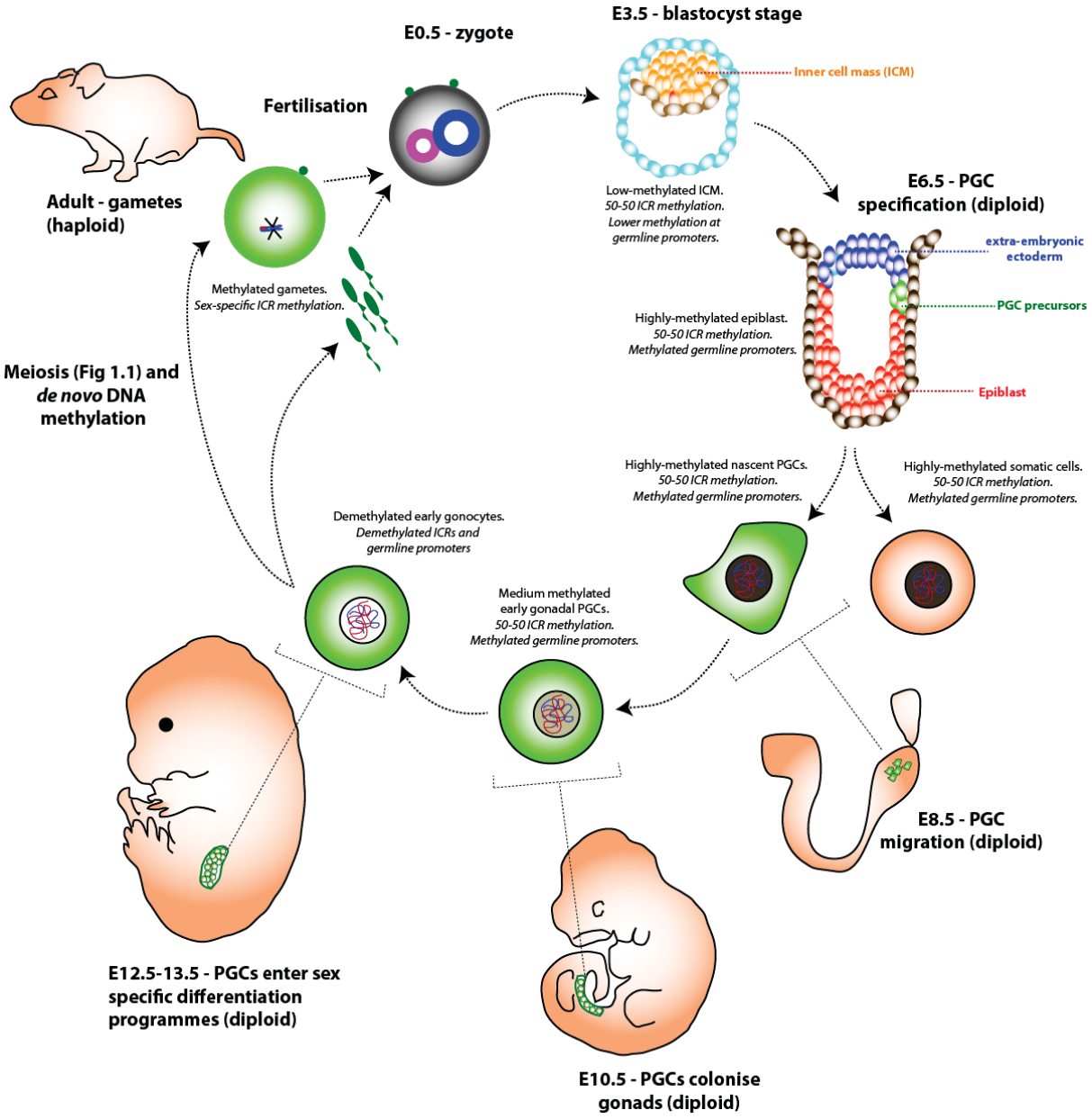


Figure 1.3 – The germ cell cycle - adapted from (Hill et al., 2014). *Key stages in germ cell development are depicted. Specific reference is made to ploidy and DNA methylation levels (globally, at imprint control regions (ICRs) and at germline gene promoters). Note that global DNA methylation levels in PGCs and early gonocytes are depicted by shading of the nucleus. Also note that "50-50 ICR methylation" refers to sex-specific mono-allelic methylation at ICRs in diploid cells, which gives typical 50% methylation level at these regions.*

While the aforementioned studies imply that germ cell specification in mouse is driven by induction from pluripotent precursor cells, gene knockout studies have since provided definitive confirmation, and have additionally identified a series of extrinsic factors that critically mediate this process,

including: the bone morphogenic protein (BMP) ligands *Bmp4*, *Bmp8b*, and *Bmp2*; the *Alk2* BMP type-1 receptor; and the BMP-signalling transducers *Smad1* and *Smad5* (Figure 1.4) (Chang and Matzuk, 2001; de Sousa Lopes et al., 2004; Lawson et al., 1999; Tremblay et al., 2001; Ying et al., 2000; Ying and Zhao, 2001). Moreover, recent comprehensive organ culture and *in vitro* germ cell differentiation experiments from Saitou and colleagues have confirmed the role of BMP signalling for induction of germ cell fate in mouse. In the former, the authors showed that BMP4 and, to a lesser extent, BMP2 are individually sufficient to induce PGCs in cultured epiblast in the absence of the anterior visceral endoderm (AVE) (Ohinata et al., 2009). In contrast, they show that BMP8b cannot act directly to induce PGCs, but when epiblasts are cultured in the presence of the AVE, addition of BMP8b is necessary for BMP4-mediated induction, thus presenting a model whereby BMP8b restricts inhibitory signals emanating from the AVE enabling BMP4/BMP2 driven PGC induction (Ohinata et al., 2009). In the latter, the authors recapitulated PGC specification from embryonic stem cells completely *in vitro*, creating what they term PGC-like cells (PGCLCs) (Hayashi et al., 2011). Specifically, PGCLC induction involves two-day differentiation of naïve ES cells into epiblast-like cells (EpiLCs) using FGF2 and activin A, and the subsequent three day induction of EpiLCs into PGCLCs using BMP4, BMP8a, EGF and LIF (Hayashi et al., 2011).

1.2.2 Transcriptional changes upon PGC specification

PGC specification is accompanied by significant changes at the transcriptional level, and a number of germ cell intrinsic factors have now been identified as critical for early mouse germ cell development. Screening for genes differentially expressed between PGCs and somatic neighbouring cells isolated at E7.0-E7.5 led to the discovery of potential candidate genes responsible for PGC specification, including *Blimp1* (Ohinata et al., 2005; Saitou et al., 2002). In *Blimp1* knockout embryos, a small cluster of less than 20 cells expressing the early PGC-specific markers stella/PGC7 and alkaline phosphatase (AP) are found at the base of the allantois at the 2-10 somite stage (Ohinata et al., 2005); however, these cells fail to migrate appropriately, and no more than one or two PGCs can be detected

after this stage (Ohinata et al., 2005). Further analysis revealed that these nascent *Blimp1*-null PGCs fail to robustly repress expression of *Hoxa1* or *Hoxb1* by the mid-bud stage, suggesting that Blimp1 is necessary to repress the somatic (i.e. mesodermal) transcriptional programme that appears to be transiently expressed in nascent PGCs (Kurimoto et al., 2008; Ohinata et al., 2005) (Figure 1.4). It should be noted that, in addition to repression of the somatic programme, PGC specification also results in up-regulation of the known pluripotency network, with Oct4 expressed throughout PGC development, Sox2 detected by E7.5, and Nanog detected from E7.75 (Kurimoto et al., 2008; Schöler et al., 1990; Yamaguchi et al., 2005; Yeom et al., 1996). Also of note, lineage tracing using *Blimp1*-Cre transgenic mice indicated that Blimp1-positive cells in the epiblast are in fact germline lineage-restricted (Ohinata et al., 2005), contradicting previous lineage tracing experiments that found no lineage-restricted PGCs at E6.5 (Lawson and Hage, 1994). However, the small number of Blimp1 positive cells at E6.5 (approximately six) makes it possible that these escaped labelling in the earlier experiments, or, alternatively, that labelled Blimp1-negative cells at this stage may still divide to give a Blimp1-positive, lineage restricted PGC and a somatic sister cell (Lawson and Hage, 1994; McLaren and Lawson, 2005; Ohinata et al., 2005).

In addition to *Blimp1*, two other genes, *Tfap2c* and *Prdm14*, appear to play a critical role in early PGC development (Figure 1.4). The putative Blimp1 target *Tfap2c* is expressed in PGCs from E7.25, and *Tfap2c* knockout results in a loss of PGCs around E8.0, marked by increased levels of some somatic differentiation markers, including *Hoxa1*, *Hoxb1* and *T* (Weber et al., 2010). Similarly, mice null for the SET-domain containing protein Prdm14 show a progressive loss of PGCs between E7.5 and E12.5, ultimately resulting in sterility (Yamaji et al., 2008). In contrast to Blimp1 and Tfap2c, which appear to predominantly be required for repression of somatic transcriptional programme (Kurimoto et al., 2008; Ohinata et al., 2005; Weber et al., 2010), Prdm14 appears to be necessary for transcriptional activation or up-regulation of the pluripotency associated genes *Sox2* and *stella/PGC7*, although the expression of both *Nanog* and *Oct4* are unaffected in *Prdm14*-knockout PGCs (Figure 1.4) (Kurimoto et al., 2008; Ohinata et al., 2005; Weber et al., 2010).

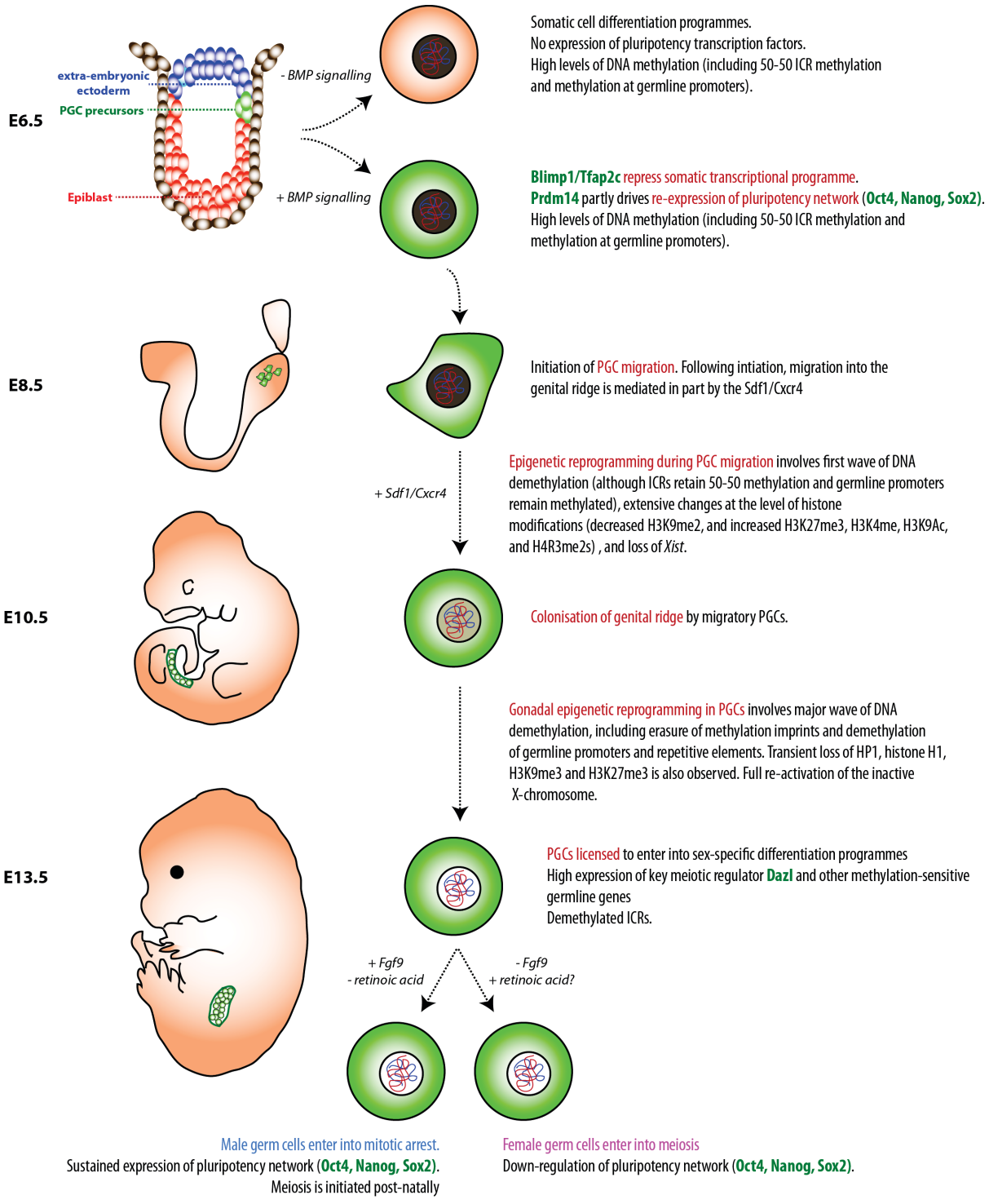


Figure 1.4 – Overview of key events in early germ cell development. *Key stages in germ cell development from PGC specification to initiation of sex specific differentiation programmes. Specific reference is made to some known extrinsic (italics) and intrinsic (bold, green) molecular regulators driving developmental changes. Note that global DNA methylation levels in PGCs and early gonocytes are depicted by shading of the nucleus. Also note that "50-50 ICR methylation" refers to sex-specific mono-allelic methylation at ICRs in diploid cells, which gives typical 50% methylation level at these regions.*

Overall, the importance of the Blimp1/Tfap2c/Prdm14 tri-partite transcription factor axis for PGC specification has been shown by both Saitou and colleagues (Nakaki et al., 2013) and Surani and colleagues (Magnúsdóttir et al., 2013). Remarkably, in the extensive study carried out by Saitou and colleagues, the authors over-expressed the three transcription factors in various combinations in epiblast-like cells (EpiLCs) to swiftly and efficiently induce cells with a transcriptional and epigenetic state reminiscent of early (i.e. E9.5) PGCs that are additionally capable of contributing to spermatogenesis and fertile offspring upon injection into seminiferous tubules of neonatal mice lacking endogenous germ cells (Nakaki et al., 2013). Surprisingly, their results revealed that Prdm14 alone, but not Blimp1 or Tfap2c, suffices for the induction of the PGC state in EpiLCs (Nakaki et al., 2013). The authors attributed this to the fact that direct expression of Prdm14 in EpiLCs bypasses the transient expression of the mesodermal program that appears to normally accompany PGC specification *in vivo*, providing further evidence that Blimp1 and Tfap2c act predominantly to repress the somatic state (Kurimoto et al., 2008; Nakaki et al., 2013; Ohinata et al., 2005; Weber et al., 2010).

1.2.3 Migration of PGCs into the genital ridge follows specification

By E7.0, PGCs are found in the posterior primitive streak in the extraembryonic mesoderm (Ginsburg et al., 1990); thereafter, they are found in the extraembryonic mesoderm at the base of the developing allantois, and, subsequently, in the underlying endoderm of the invaginating hind-gut (Ginsburg et al., 1990). From E8.5, PGCs initiate their active migration into the genital ridge, the future site of the gonads (Figures 1.3, 1.4) (Anderson et al., 2000). Migration appears to be regulated by a number of distinct signalling pathways, with gene knockout studies implicating Sdf1 (expressed at the site of the future gonads and in the surrounding mesenchyme) and its receptor Cxcr4 (expressed in PGCs) as a chemoattractant system responsible for late stage migration of PGCs into the genital ridge (Figure 1.4) (Ara et al., 2003; Molyneaux et al., 2003). In the absence of either protein, very few PGCs reach the genital ridge; moreover, the ectopic expression of Sdf1 causes PGCs to migrate to new locations (Ara

et al., 2003; Molyneaux et al., 2003). In addition to Sdf1/Cxcr4 signalling, the *steel* ligand/*c-kit* receptor pathway appears to also play an important role during PGC migration (Gu et al., 2009). However, in contrast to Sdf1/Cxcr4, gene knockout studies have suggested that steel and c-kit generally regulate PGC motility, with steel loss of function resulting in PGCs that migrate in the proper direction but at a greatly reduced rate (Figure 1.4) (Gu et al., 2009).

1.2.4 Initiation of sex-specific gonocyte differentiation programmes in mouse PGCs

Immediately following colonisation of the gonads, male and female murine primordial germ cells are bi-potential, equally capably of progressing through the female or male differentiation programmes, as clearly shown in two elegant studies by McLaren and colleagues. In the first, the authors isolated germ cells from female and male E11.5 embryos, and observed meiotic initiation and progression through prophase I for both XX and XY PGCs when re-aggregated with disaggregated embryonic lung cells (McLaren and Southee, 1997) or, alternatively, E12.5 female urogenital ridges (Adams and McLaren, 2002). However, they also observed that the potential for male germ cells to enter into meiosis is lost by E12.5 (Adams and McLaren, 2002; McLaren and Southee, 1997). In the second study, the authors studied the inverse relationship, re-aggregating isolated E11.5 female PGCs with disaggregated E12.5 urogenital ridges from male embryos, and clearly observed development of XX PGCs to spermatogonia (Adams and McLaren, 2002). Again however, the bi-potentiality of female PGCs is lost by E13.5, although unlike in males, it is still observed at E12.5 (Adams and McLaren, 2002).

While E11.5 PGCs are equally capable of progression through male and female development, normally germ cells start to undergo sex-specification at this point, initiating their sex specific gonocyte differentiation programmes (Figure 1.4). Sex specification in mouse is driven by germ cell

extrinsic rather than intrinsic factors, with multiple signalling molecules inducing different aspects of sex-specific germ cell behaviour in male and female embryos. In males, transient expression of the Y-chromosome linked gene *Sry* in the somatic cells of the gonad starting at E10.5 results in up-regulation of the transcription factor *Sox9*, thus ultimately inducing differentiation of these supporting cells into male Sertoli cells rather than female granulosa cells (Koopman et al., 1991; Sekido et al., 2004; Sekido and Lovell-Badge, 2008). Sertoli cells then signal to and masculinize other cell types in the gonad and adjacent mesonephros, including germ cells, to induce testis differentiation (Kocer et al., 2009). Gene knockout and *in vitro* culture experiments have revealed that masculinisation of germ cells appears to be mediated by expression of two proteins expressed by Sertoli cells, the Fgf9 signalling ligand (Figure 1.4), which can act directly on germ cells, and the P450 enzyme Cyp26b1, which acts indirectly on germ cells, potentially by degrading retinoic acid released from the mesonephros and thus preventing retinoic acid signalling in XY gonocytes (Bowles et al., 2010; Bowles et al., 2006; Colvin et al., 2001; Kim et al., 2006; MacLean et al., 2007). It should be noted that, in addition to Fgf9 and Cyp26b1, other signalling components expressed in somatic lineages have also been shown to play important roles in male germ cell specification, including the membrane trafficking protein Sdmg1 (Best et al., 2008). Ultimately, signalling from Sertoli cells induces expression of the essential male germ cell factor Nanos2 in E13.5 XY germ cells (Suzuki and Saga, 2008; Tsuda et al., 2003); prevents meiotic entry (Bowles et al., 2010; Bowles et al., 2006; Gill et al., 2011; Kocer et al., 2009); enables the continued expression of the pluripotency transcription factors (Bowles et al., 2010); and ultimately allows for progression through the male foetal gonocyte differentiation programme (Figure 1.4) (Bowles et al., 2010; Suzuki and Saga, 2008). It should be noted that male germ cells enter into meiosis only postnatally, at which point they will complete their gametic differentiation (Figures 1.1, 1.4) (Kocer et al., 2009).

In female embryos lacking the Y chromosome, *Sry* expression does not occur, and consequently the gonadal somatic cell lineage differentiates into female granulosa cells rather than male Sertoli cells (Kocer et al., 2009); this process additionally appears to be mediated by somatically expressed Wnt

signalling molecules such as Rspo1 and Wnt4, which act by down-regulating *Sox9* expression in the supporting cell lineage (Chassot et al., 2008; Kim et al., 2006; Vainio et al., 1999). In the absence of Sertoli cells, Cy26b1 is not expressed, and thus E12.5 XX germ cells are exposed to RA signalling released from the mesonephros (Figure 1.4) (Bowles et al., 2010; Bowles et al., 2006). In turn, a number of groups have shown that exposing both XX or XY E12.5 gonads to non-physiologically high levels of retinoic acid in (serum-free) organ culture experiments leads to the transcriptional up-regulation of a number of key meiotic regulators, including *Stra8*, *Sycp3*, *Dmc1* and *Spo11*, thus suggesting that retinoic acid may act as a meiosis initiating factor driving female foetal gonocyte differentiation (Anderson et al., 2008; Bowles et al., 2006; Lin, 2008). However, it should be noted that the importance of retinoic acid for meiosis initiation in female foetal germ cells has been the subject of intense debate. Specifically, experiments performed in a serum-containing system found that the vast majority of E11.5 XY germ cells cultured with RA or Cyp26b1 inhibitor remained as prospermatogonia (Best et al., 2008). Moreover, using *Raldh2*$^{-/-}$ embryos, Duester and colleagues showed that *Stra8* is expressed in PGCs in the absence of physiologically detectable levels of RA, although the authors of this study also note that inhibition of Cyp26b1 in *Raldh2*$^{-/-}$ testis allows RA-independent induction of *Stra8* only when the mesonephros remains attached (Kumar et al., 2011). This suggests a non-RA signal from the mesonephros induces *Stra8* in the adjacent gonad (Kumar et al., 2011). Nonetheless, in contrast to males, female gonocytes initiate meiotic prophase starting at E13.5, concurrent with down-regulation of the major pluripotency network (Figure 1.4) (Seisenberger et al., 2012). It should be noted, however, that oocytes remain arrested in meiotic prophase from before birth until post-puberty, when a surge of gonadotropins triggeres meiotic progression to resume. Ultimately, meiosis will only complete following the initiation of fertilisation (Figure 1.1).

1.3 Epigenetics and the acquisition of the unique properties of murine gametes

In contrast to somatic cell differentiation, where gene-expression programmes become more restricted through the deposition of complex layers of epigenetic modifications (Hill et al., 2014), once germ cell fate has been specified, there is no need for further developmental restriction: PGCs will give rise to gametes, and those that fail to do so will die. Nonetheless, the epigenetic changes observed in both the embryonic and post-natal precursors of the gametes are necessary for the correct progression of these cells through the germ cell developmental programme. For simplicity, the epigenetic changes observed during this process can broadly be divided into an initial reprogramming phase during which the epigenetic signature associated with epiblast precursor cells is lost and a programming phase during which the sex-specific gamete epigenetic template is established.

1.3.1 Overview of epigenetic reprogramming in mouse primordial germ cells

1.3.1.1 Epigenetic reprogramming during PGC migration

Epigenetic reprogramming in primordial germ cells is characterised by two stages. The first stage occurs during migration of nascent PGCs from their location in the underlying endoderm of the invaginating hindgut into the genital ridge. Global changes with respect to histone modifications have been extensively studied *in vivo* by immunofluorescence. These studies have revealed near complete loss of H3K9me2 following E8.0, concurrent with increased H3K27me3, H3K4me and H3K9Ac, and H4R3me2s (Figure 1.4) (Ancelin et al., 2006; Hajkova et al., 2008; Seki et al., 2005; Seki et al., 2007). Additionally, early changes at the level of histone modifications have been investigated in detail at the locus specific level by ChIP-Seq for H3K27me3, H3K4me3, H3K27Ac, and H3K9me2 in PGC-like cells (PGCLCs), which re-capitulate early (pre-E11.5) PGC development (see Section 1.2.1

for details) (Kurimoto et al., 2015). These studies corroborated the immunofluorescence data, and additionally showed that PGCLCs initially lose H3K4me3 from many bivalent genes at the time of specification only to regain this mark with concomitant increase in H3K27me3 levels, and that the progressive loss of H3K9me2 occurs at lamina-associated perinuclear heterochromatin, which, along with reduced concentration of LAMIN B1 around the nuclear periphery, results in changes in nuclear architecture (Hajkova et al., 2008; Kurimoto et al., 2015).

Beyond the extensive changes at the level of histone modifications, a first wave of DNA demethylation occurs during migration, as E9.5 PGCs have a globally hypo-methylated state compared with the highly methylated epiblast cells from which they are originally specified (Figure 1.4) (Guibert et al., 2012; Seisenberger et al., 2012). This involves roughly a 50% reduction of methylation at promoters, gene bodies, and intergenic regions, although significant methylation is still retained at retrotransposons, imprint control regions and germline gene promoters (Figures 1.3, 1.4) (Seisenberger et al., 2012).

1.3.1.2 Gonadal epigenetic reprogramming

Following migration into the genital ridge, primordial germ cells undergo a second phase of epigenetic reprogramming, henceforth referred to as gonadal epigenetic reprogramming (Figure 1.4). This second phase was originally characterised by erasure of methylation imprints (Figures 1.3, 1.4, 1.5A-B), as revealed by locus specific bisulphite-sequencing from both PGCs directly isolated from mid-gestation embryos and from somatic cell nuclear transfer (SCNT) embryos obtained using gonadal PGC donor nuclei (Hajkova et al., 2002; Lee et al., 2002). However, further investigation using various genome-wide next generation sequencing approaches (including MeDIP, reduced representation bisulphite sequencing, and whole genome bisulphite sequencing), revealed extensive

global DNA demethylation beyond simple imprint erasure, with the bulk of methylation additionally lost at most regulatory regions, including germline promoters, and repetitive elements, including retrotransposons (Figures 1.3, 1.4, 1.5C) (Guibert et al., 2012; Seisenberger et al., 2012).

Concurrent with gonadal DNA demethylation is the transient loss of heterochromatin associated protein HP1, histone H1, and H3K9me3 and H3K27me3 marks, as determined by immunofluorescence (Figure 1.4) (Hajkova et al., 2008), and reactivation of the inactive X-chromosome, as determined by an elegant flow-cytometry based strategy using transgenic PGCs expressing GFP from the X-chromosome (Figure 1.4) (Chuva de Sousa Lopes et al., 2008). It should be noted that while expression from the inactive X-chromosome only appears to be initiated upon entry of PGCs into the genital ridge (Chuva de Sousa Lopes et al., 2008), many studies have confirmed that the process is initiated during migration, with loss of *Xist* expression already observed in nascent PGCs (Figure 1.4) (de Napoles et al., 2007; Sugimoto and Abe, 2007).

1.3.2 The importance of epigenetic reprogramming for germ cell development

1.3.2.1 Epigenetic reprogramming and licensing for gametogenesis

As discussed Section 1.2.4, cell extrinsic factors play an essential role in triggering spermatogenesis (Fgf9 signalling, lack of retinoic acid), and potentially oogenesis (retinoic acid signalling (?)). However, these alone cannot explain the temporal and cell-specific nature of the female and male differentiation programmes as somatic cells exposed to these extrinsic cues do not initiate gametogenesis (Lesch and Page, 2012). Rather, the specific expression of germ cell intrinsic factors is additionally necessary (Gill et al., 2011; Lin, 2008), with the translation regulator Dazl representing the most well studied of these factors (Figures 1.4, 1.5C). Dazl was originally identified as a germ cell

specific RNA-binding protein in mammals, the expression of which is first detected in E12.5 PGCs (Saxena et al., 1996; Seligman and Page, 1998). Gene knockout studies subsequently revealed that Dazl is critical for the expression of core regulators of both male and female gonocyte differentiation programmes (Gill et al., 2011; Lin, 2008), and in the absence of Dazl, both male and females are completely sterile (Ruggiu et al., 1997). More specifically, *Dazl*-knockout germ cells isolated from inbred embryos are unable to express the key meiotic regulator *Stra8,* which is required for premeiotic DNA replication, in foetal ovaries in response to retinoic acid (Figure 1.5C) (Lin, 2008); thus female gonocytes fail to initiate meiotic recombination, ultimately leading to germ cell depletion by E19.5 (Lin, 2008; Ruggiu et al., 1997). Similarly, male germ cells isolated from inbred embryos deficient of Dazl are unable to express *Nanos2*, fail both to arrest in G1/G0 and to remethylate their genomes, and undergo extensive apoptosis starting at E15.5 (Figure 1.5C) (Gill et al., 2011; Lin and Page, 2005). It should be noted that the timing of male germ cell loss in *Dazl*-knockout mice appears to vary depending on the genetic background, with the loss of male germ cells occurring later on an outbred background (reviewed in (Brook et al., 2009)). Morever, the final block in gametogenesis in germ cells isolated from outbred mice deficient of Dazl has been shown by careful immunological analysis to be a failure to progress beyond leptotene of meiotic prophase I, and not a failure in premeiotic DNA replication due to impaired Stra8 expression (Saunders et al., 2003).

At the molecular level, *Dazl* is a member of a group of approximately 200 genes whose transcription, and thus expression, is highly sensitive to the methylation states of their promoter regions (Figure 1.5C) (Karimi et al., 2011; Tsumura et al., 2006). This classification is based on the significantly increased mRNA levels for these genes in *Dnmt1*/*Dnmt3a*/*Dnmt3b*-triple knockout (Dnmt-TKO) embryonic stem cells (Karimi et al., 2011), their premature transcription *in vivo* in *Dnmt1*-mutant embryos largely devoid of methylation (Maatouk et al., 2006), and the restriction of their normal expression to the hypo-methylated cells of the inner cell mass and gonadal PGCs following DNA demethylation (Blaschke et al., 2013; Seisenberger et al., 2012). Consistent with their germ cell specific expression pattern, gene ontology (GO) analysis revealed that “regulation of meiotic cell

cycle", "oogenesis" and "reproductive cellular processes" terms are significantly enriched among methylation-sensitive genes, including the key meiosis regulators *Dazl*, *Stra8*, *Hormad1* and *Hormad2*, among others (Karimi et al., 2011). In view of these collective results, the transcription of methylation-sensitive genes appears to require promoter demethylation (Figure 1.5C), and thus the wave of global DNA demethylation in PGCs represents a potentially critical step in licensing germ cells for gametogenesis by enabling expression of germ cell intrinsic licensing factors such as Dazl.

Beyond DNA demethylation, changes with respect to histone modifications have recently emerged as additional regulators of gonocyte-specific gene expression. Specifically, ChIP experiments have implicated a role for both polycomb repressive complex (PRC) 1 and PRC2, respectively responsible for deposition of H2AK119u1 and H3K27me3 histone modifications, in repressing expression of gonocyte specific genes (e.g. *Dazl*) until E12.5 in both sexes, and meiosis specific genes (e.g. *Stra8*) until E12.5 in females (Lesch et al., 2013; Ng et al., 2013; Sachs et al., 2013; Yokobayashi et al., 2013). Additionally, conditional gene knockout of PRC1 components *Rnf2* and *Ring1* revealed that loss of PRC1 activity at E11.5 induces precocious induction of meiosis in females due to the early expression of meiosis-specific genes (Yokobayashi et al., 2013). Intriguingly, the authors of this study also noted that precocious initiation of meiosis was associated with an inability to complete the meiotic programme and thus to develop into mature oocytes (Yokobayashi et al., 2013), suggesting that regulation of meiotic timing by chromatin modifications may be essential for successful gametogenesis.

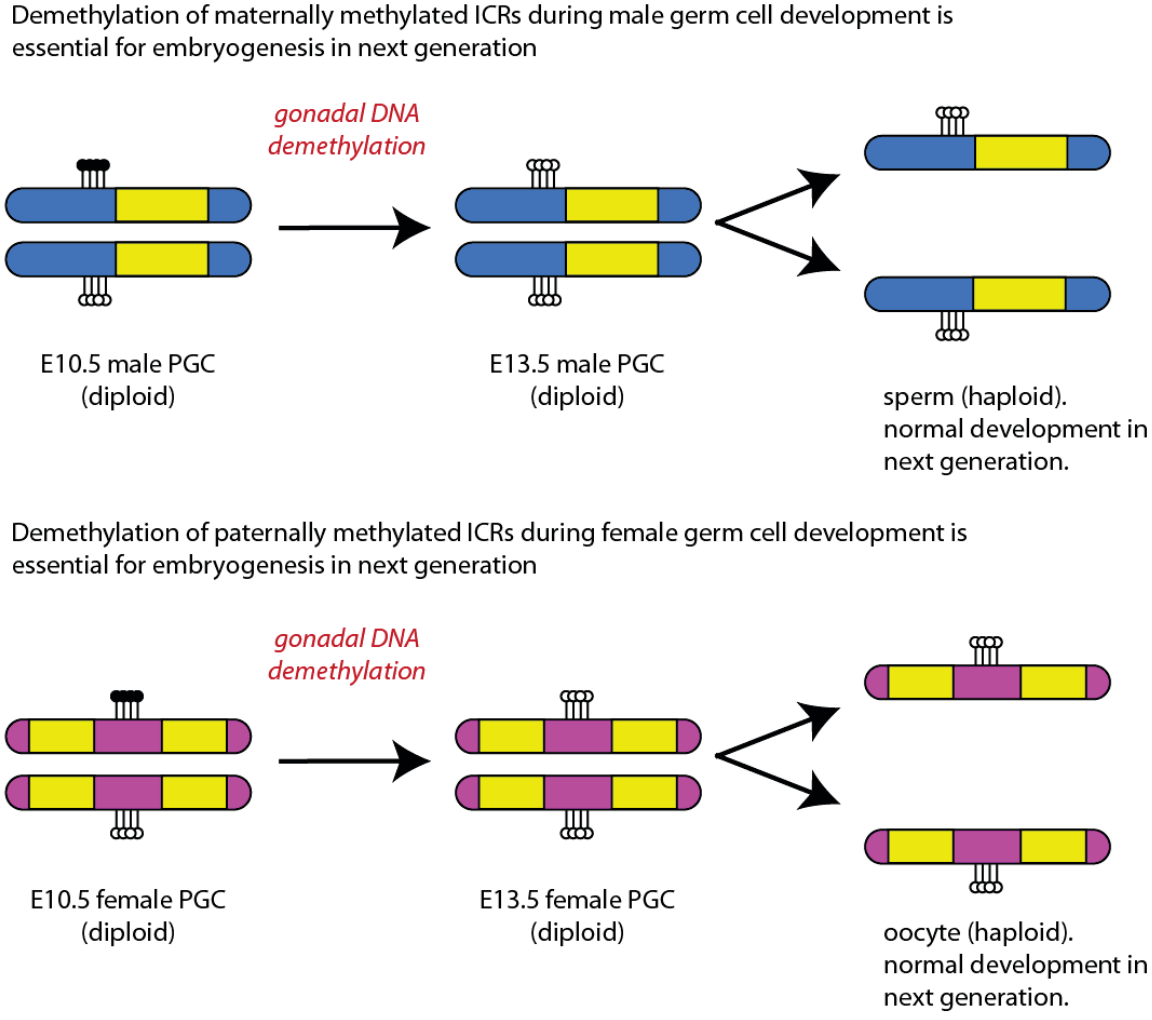


Figure 1.5 – The importance of gonadal DNA demethylation for germ cell development. A-B) *In mouse, DNA demethylation of imprint control regions (ICRs) that are normally unmethylated in either male (A) or female (B) gametes represents a critical event during male and female germ cell development, respectively. In theory, germ cells that fail to demethylate these regions will give rise to a high proportion of gametes that have methylation patterns incompatible with post-implantation embryonic development in the next generation.* ***C)*** *The transcription of methylation-sensitive germline*

genes, such as the translation-regulator Dazl, appears to require promoter demethylation, and expression of these methylation sensitive germline genes is essential for continued progression through both male and female gonocyte differentiation programmes, including meiosis.

1.3.2.2 Imprint erasure and the correct establishment of methylation imprinting in gametes

As detailed in section 1.1.2, the establishment of the correct sex-specific methylation patterns at imprint control regions (ICRs) in gametes is essential for embryonic development in the next generation. However, ICRs are particularly refractory to changes in their methylation state, as evidenced by the persistence of mono-allelic methylation in all tissues except the germ line from the moment of fertilisation into adult life, including during the global wave of DNA demethylation observed in early embryogenesis (Figure 1.3) (Howell et al., 2001; Li et al., 1993; Li et al., 2008; Nakamura et al., 2007). Nevertheless, DNA demethylation of imprint control regions (ICRs) that are normally unmethylated in either male or female gametes represents a critical event during male and female germ cell development, respectively (Figures 1.5A-B). In theory, germ cells that fail to demethylate these regions will give rise to a high proportion of gametes that have methylation patterns incompatible with post-implantation embryonic development in the next generation (Figure 1.2C). Of note, in the reciprocal event, in which ICRs that are normally methylated in either male or female gametes fail to re-methylate during male and female germ cell development, respectively, a similar severe fertility phenotype is observed (Figure 1.2D), as evidenced by consistent embryonic death of the heterozygous offspring of females with oocytes depleted of Dnmt3a or Dnmt3L, the enzymes responsible for *de novo* methylation of ICRs during oogenesis (see section 1.1.2 for further details) (Bourc'his et al., 2001; Kaneda et al., 2004).

1.4 Developmentally-associated global DNA demethylation events in mammals

As discussed in section 1.3, DNA demethylation in gonadal PGCs appears to play a critical role for both licensing primordial germ cells for gametogenesis and for the correct establishment of sex specific methylation patterns at imprint control regions in gametes. As such, research into the molecular triggers and mechanisms driving DNA demethylation in gonadal PGCs has been an area of intense investigation over the past decade. Global DNA demethylation, however, is not unique to murine gonadal PGCs. Rather, waves of global DNA demethylation (and subsequent re-methylation) are observed during germ cell development in other mammalian species, including humans (Gkountela et al., 2015; Guo et al., 2015; Hyldig et al., 2011; Tang et al., 2015). Moreover, (more limited) global DNA demethylation and re-methylation is also observed during pre- and early post-implantation development in all mammalian species studied to date (Amouroux et al., 2016; Gu et al., 2011; Guo et al., 2014a; Guo et al., 2014b; Mayer et al., 2000; Oswald et al., 2000; Shen et al., 2014; Smith et al., 2014; Wang et al., 2014; Wossidlo et al., 2011), as well as during the (reversible) transition between naive and primed pluripotency in embryonic stem cells cultured *in vitro* (Ficz et al., 2013; Habibi et al., 2013; Leitch et al., 2013; Takashima et al., 2014). Thus, understanding the mechanisms underlying global DNA demethylation represents a central question in mammalian developmental biology.

In this section, I first broadly introduce the potential mechanisms of DNA demethylation in mammals. I next discuss in more detail our current understanding of the mechanisms of global DNA demethylation (and re-methylation) in the mouse embryo, with a particular emphasis on the DNA demethylation events observed in mouse PGCs.

1.4.1 Potential mechanisms of DNA demethylation

1.4.1.1 Passive vs. active DNA demethylation

Conceptually, DNA demethylation can be classified as either occurring via a passive or active mechanism depending on its requirement for DNA replication (Figure 1.6) (reviewed in (Hill et al., 2014)). During a typical cell division, DNA methylation patterns on the parent DNA strands are faithfully copied onto the newly synthesized daughter strands by the maintenance DNA methyltransferase Dnmt1, which is recruited in a PCNA- and Uhrf1-dependent manner to the replication fork, and which specifically methylates hemi-methylated DNA (Figure 1.6A) (Bostick et al., 2007; Chuang et al., 1997; Li et al., 1992; Sharif et al., 2007). In the absence of either Dnmt1 or Uhrf1, DNA methylation is passively lost over the course of a number of cell divisions, ultimately resulting in a demethylated cell population (Figure 1.6A) (Bostick et al., 2007; Li et al., 1992; Sharif et al., 2007).

Alternatively, DNA methylation can be removed via an active enzymatic mechanism independent of a requirement for DNA replication, followed by replacement of the excised site by activation of the DNA repair machinery (Figure 1.6B) (Hill et al., 2014), as has been well characterised in plants *in vivo* (Morales-Ruiz et al., 2006) and additionally suggested in mammals *in vitro* (Barreto et al., 2007). Intriguingly, however, the family of 5mC specific DNA glycosylases clearly responsible for active DNA demethylation in flowering plants (Morales-Ruiz et al., 2006) appears to have no direct sequence homology in higher vertebrates, including mouse, and thus active DNA demethylation via enzymatic removal of 5mC in these organisms may require additional modification of the base to trigger the DNA repair response (Figure 1.6B) (Hill et al., 2014).

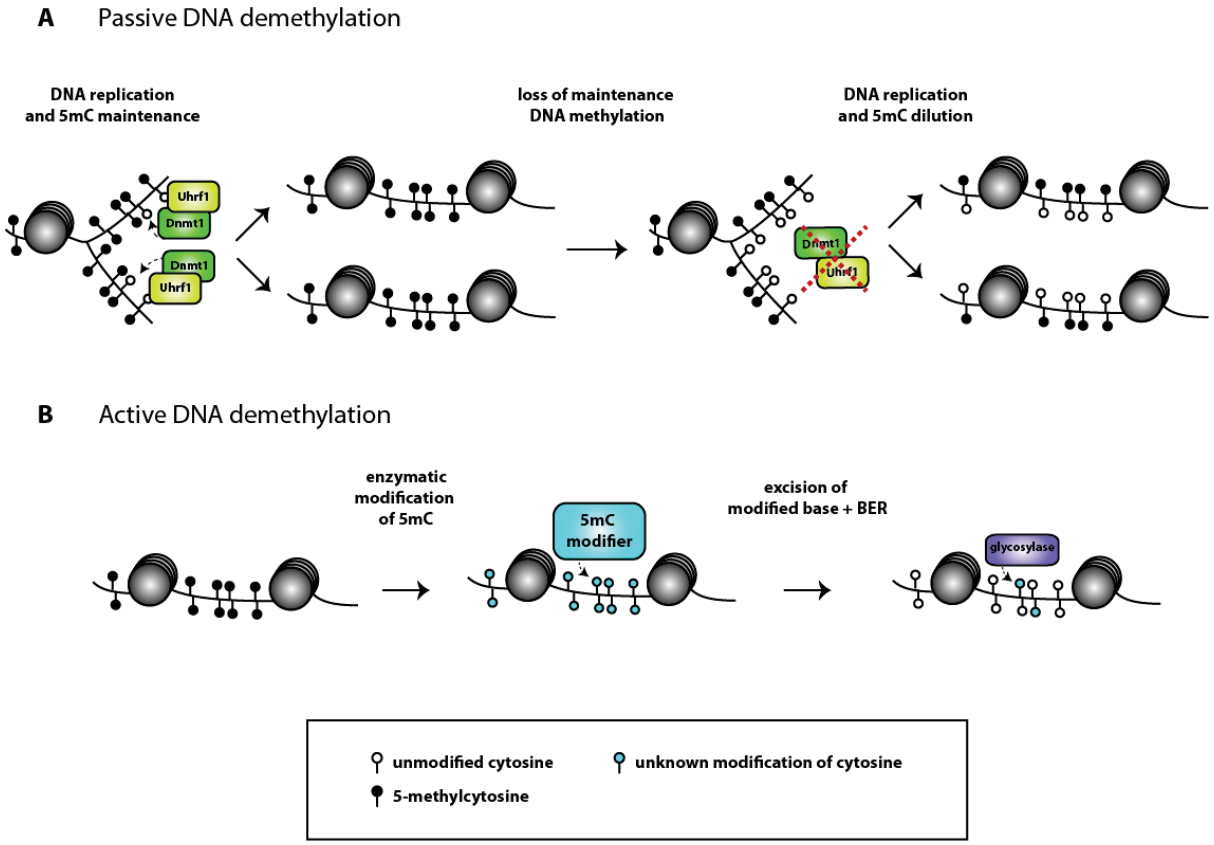


Figure 1.6 – Potential mechanisms of DNA demethylation in mouse - adapted from (Hill et al., 2014). A) *Passive DNA demethylation involves a lack of methylation maintenance during DNA replication, leading to dilution of 5mC over time and, eventually, fully unmethylated DNA.* ***B)*** *Active DNA demethylation in mouse likely involves enzymatic modification of 5mC, followed by excision of the modified base by a DNA glycosylase and replacement of the abasic site by base excision repair (BER). Aid and Tet proteins have been proposed as 5mC modifying enzymes during active DNA demethylation (Hackett et al., 2013; Hajkova et al., 2010; Popp et al., 2010; Yamaguchi et al., 2012; Yamaguchi et al., 2013b).*

1.4.1.2 Tet proteins and the enzymatic modification of 5mC

Recently, the *Ten-eleven translocation* (Tet) family of 5mC oxygenases, consisting of Tet1, Tet2 and Tet3, have come to the forefront of the mammalian DNA (de)methylation field due to their ability to oxidise 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and its higher oxidative derivatives (Tahiliani et al., 2009), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC; Figure1.7) (He et al., 2011). Tet1, the first of the Tet proteins discovered, was originally identified as a fusion partner of the H3K4 methyltransferase MLL in acute myeloid leukaemia (Lorsbach et al., 2003; Ono

et al., 2002). Despite its unremarkable discovery, elegant work by Rao and colleagues revealed in 2009 that Tet1 is a 5mC oxygenase in mammals (Figure 1.7) (Tahiliani et al., 2009). This finding was based on the homology between Tet1 and the trypanosome thymine oxygenases JBP1 and JBP2, as well as the appearance of 5-hydroxymethylcytosine (5hmC) in DNA upon overexpression of Tet1 full-length protein or the Tet1 catalytic domain in cultured mammalian cells (Tahiliani et al., 2009). It has subsequently been shown that the Tet1 homologs Tet2 and Tet3 are also capable of oxidising 5mC (Figure 1.7) (Ito et al., 2010), albeit with lower efficiency (Fu et al., 2014), and that 5hmC, 5fC and 5caC are physiological constituents of mammalian DNA both *in vitro* and *in vivo*, although their absolute levels are consistently an order of magnitude (5hmC) or two (5fC, 5caC) lower than those observed for 5mC (Bachman et al., 2015; Bachman et al., 2014; Kriaucionis and Heintz, 2009).

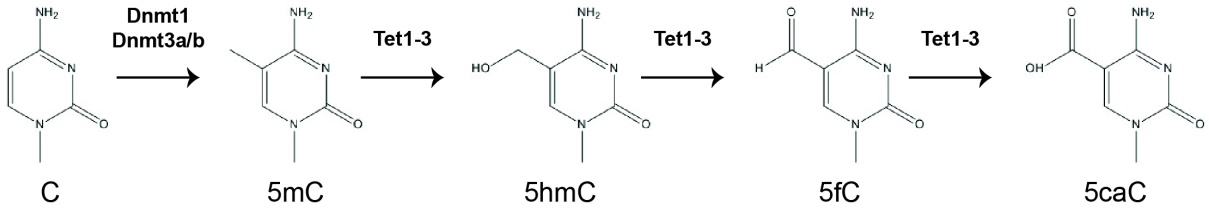


Figure 1.7 – Overview of Tet protein enzymatic activity in mouse - adapted from (Hill et al., 2014).

Since the initial discovery of Tet enzymatic activity, intense research has focussed on elucidating the biological significance of Tet-driven 5mC oxidation, with early biochemical evidence suggesting a possible mechanistic role in both passive and active DNA demethylation (reviewed in (Hill et al., 2014)). With respect to the former, the maintenance methyltransferase Dnmt1, which shows a preference for hemi-methylated DNA and is recruited to the replication fork where it normally faithfully copies DNA methylation patterns of the parent DNA strands onto newly synthesized daughter strands (Figure 1.8A) (Bostick et al., 2007; Chuang et al., 1997; Li et al., 1992; Sharif et al., 2007), has been shown to have low enzymatic activity on hemi-hydroxymethylated DNA (Figure 1.8A) (Valinluck and Sowers, 2007). Thus, following oxidation of 5mC to 5hmC by Tet proteins,

DNA methylation can be gradually lost over rounds of DNA replication and cell division (Figure 3.2A) (Hill et al., 2014). With respect to active DNA demethylation, biochemical studies by Xu and colleagues showed that 5fC and 5caC can be excised by thymine-DNA glycosylase (TDG), and that, upon depletion of TDG, embryonic stem cells accumulate 5caC to readily detectable levels (He et al., 2011). Oxidation of 5mC by Tet proteins thus represents a critical first step in active TDG-mediated DNA demethylation (Figure 1.8B) (He et al., 2011)

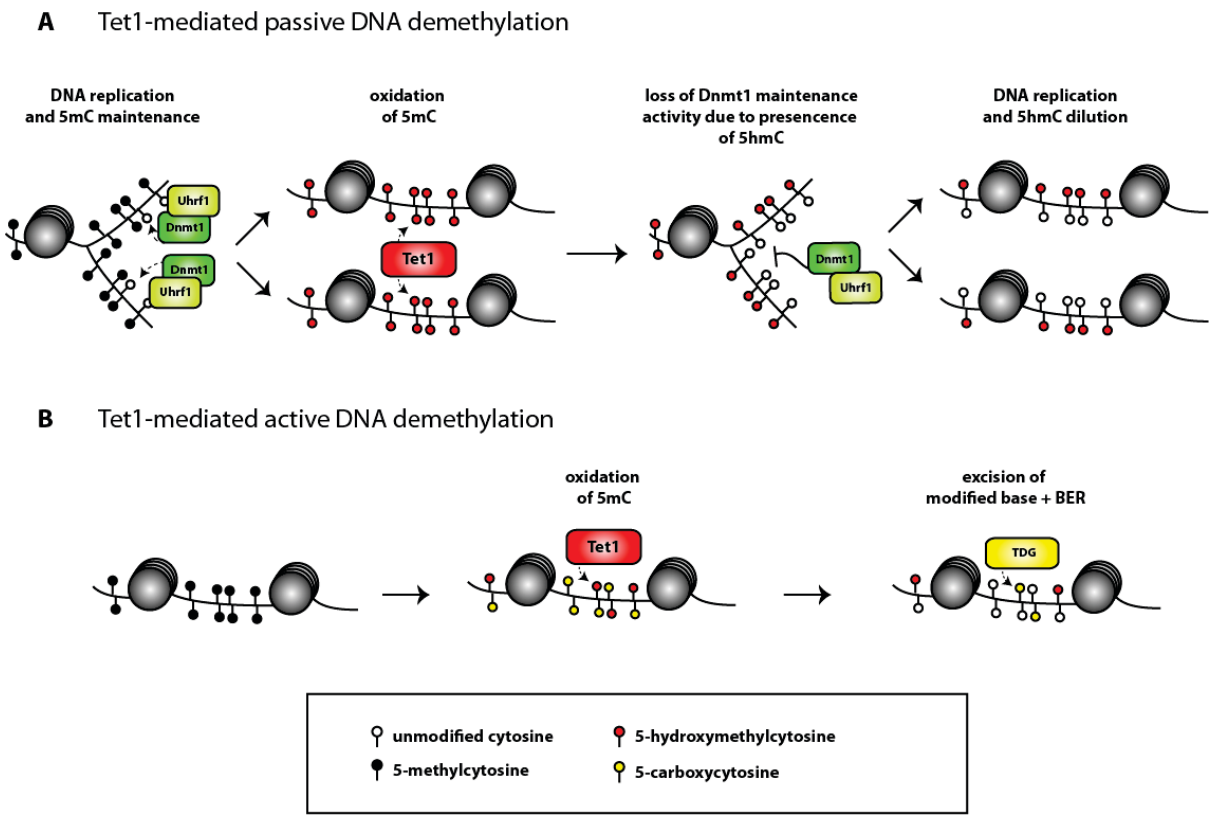


Figure 1.8 – Potential mechanisms of Tet1-mediated DNA demethylation - adapted from (Hill et al., 2014). A) *Tet1-mediated passive DNA demethylation first involves oxidation of 5mC to 5hmC. As Dnmt1 has been shown to have lower activity on hemi-hydroxymethylated DNA, a lack of methylation maintenance during DNA replication can lead to dilution of 5mC over time and, eventually, to fully unmethylated DNA.* ***B)*** *Tet1-mediated active DNA demethylation in mouse first involves oxidation of 5mC to 5caC via 5hmC, followed by excision of 5caC by thymine DNA glycosylase (TDG) and replacement of the abasic site by base excision repair (BER)..*

1.4.2 The DNA demethylation/re-methylation cycle in mammals

1.4.2.1 DNA demethylation and re-methylation dynamics in pre- and early post-implantation mammalian embryos

The first instance of global DNA demethylation observed in mouse occurs immediately following fertilisation when the highly condensed and methylated paternal genome is first decondensed (through protamine removal and replacement with the histone variant H3.3) and subsequently demethylated (Amouroux et al., 2016; Mayer et al., 2000; Oswald et al., 2000; Torres-Padilla et al., 2006; Wang et al., 2014). DNA demethylation on the paternal genome occurs at most genic features and repetitive elements, although imprint control regions and IAP elements are largely unaffected (Wang et al., 2014). Of note, similar results have also been observed in humans (Guo et al., 2014b; Smith et al., 2014).

Numerous reports using aphidicolin treatment, which inhibits DNA synthesis, during *in vitro* fertilisation (IVF) coupled with immunofluorescence (IF) or liquid chromatography-mass spectrometry (LC/MS) analysis have shown that the bulk of DNA methylation on the paternal genome is lost a few hours following fertilisation via a rapid replication-independent mechanism (Amouroux et al., 2016; Hajkova et al., 2010; Mayer et al., 2000; Oswald et al., 2000). Consistently, high levels of the Base Excision Repair (BER) components XRCC1, Ape1, and Parp1 are detected and chromatin bound specifically on the paternal genome during zygotic DNA demethylation (Hajkova et al., 2010), with significantly increased methylation levels detected on the paternal DNA of zygotes obtained by *in vitro* fertilisation with media containing 3-AB, a Parp1 inhibitor which delays progression of BER (Hajkova et al., 2010). It should be noted that, although the bulk of DNA demethylation occurs via a rapid replication-independent mechanism, a small subset of loci appears to require DNA replication to complete DNA demethylation (Guo et al., 2014a; Shen et al., 2014).

In agreement with the possible involvement of 5hmC in DNA demethylation (see section 1.4.1 for details), Tet3 oxygenase has been shown to be highly expressed in mouse oocytes (Amouroux et al., 2016; Gu et al., 2011; Wossidlo et al., 2011). Additionally, IF staining using 5hmC specific antibodies identified this modification enriched in the paternal pronucleus following DNA demethylation, with clear 5hmC signal still detectable in half of the nucleus corresponding to paternal chromatin in the two-cell stage (Amouroux et al., 2016; Gu et al., 2011; Inoue and Zhang, 2011; Wossidlo et al., 2011). Considering these observations, and that *Tet3*-deficient oocytes, which are unable to catalyse formation of 5hmC during zygotic DNA demethylation, showed increased 5mC levels globally on the paternal pronucleus (Gu et al., 2011; Wossidlo et al., 2011), it was originally hypothesized that zygotic DNA demethylation proceeds through oxidation of 5mC to 5hmC, which is followed by passive loss of 5hmC through DNA replication (Gu et al., 2011; Inoue and Zhang, 2011; Wossidlo et al., 2011). However, this model is inconsistent with the aforementioned experiments using aphidicolin during IVF, which strongly suggest that the bulk of DNA demethylation proceeds via an active process (Amouroux et al., 2016; Hajkova et al., 2010; Mayer et al., 2000; Oswald et al., 2000). Moreover, more recent studies using LC/MS and whole genome bisulphite sequencing to assess the levels of 5mC in late stage (i.e. PN4-5) zygotes maternally depleted of Tet3 have since shown that only limited hyper-methylation is observed in the absence of 5mC oxidation (Amouroux et al., 2016; Peat et al., 2014).

These apparent inconsistencies were elegantly explained in a recent study by Hajkova and colleagues in which the authors studied in detail the kinetics of DNA demethylation and 5mC oxidation on the paternal pronucleus (Amouroux et al., 2016). The authors observed that the paternal genome is first actively demethylated before Tet3-mediated 5hmC appears; in other words, 5mC disappearance and 5hmC appearance in the paternal genome are uncoupled (Amouroux et al., 2016). Following the bulk of DNA demethylation, the authors showed that new 5mC is deposited via the activity of the de novo

DNA methyltransferase Dnmt3a and maintenance DNA methyltransferase Dnmt1, previously thought to be inactive during zygotic epigenetic reprogramming (Amouroux et al., 2016). Finally, the authors show that it is this newly formed 5mC which is targeted by zygotic Tet3, preventing re-accumulation of DNA methylation (Amouroux et al., 2016). It should be noted, however, that the targets and overall biological significance of this two-tier demethylation process remain to be established.

From the two-cell stage onwards, both the murine paternal and maternal genomes undergo progressive loss of methylation until the blastocyst stage, when the embryonic genome contains very low levels of DNA methylation just prior to implantation (Okamoto et al., 2016; Wang et al., 2014). As with zygotic DNA demethylation, most genic features and repetitive elements undergo DNA demethylation between the two-cell and blastocyst stage embryo, although imprint control regions and IAP elements are again largely unaffected (Wang et al., 2014). It should be noted that, to date, little is known regarding the mechanism underlying this wave of DNA demethylation; however, 5mC kinetics as assessed by both whole genome bisulphite sequencing and LC/MS would be consistent with a model of passive dilution (Okamoto et al., 2016; Wang et al., 2014). Intriguingly, in humans, the vast majority of DNA demethylation appears to have already occurred by the two-cell stage, and thus only limited further DNA demethylation is seen between two-cell stage and blastocyst stage (Guo et al., 2014b). However, in both mouse and humans, high levels of DNA methylation are subsequently re-established at most genic features and repetitive elements in the epiblast immediately following implantation, from which primordial germ cells are eventually specified (Guo et al., 2014b; Wang et al., 2014). In mouse, re-methylation is mediated by the specific expression of Dnmt3b *de novo* DNA methyltransferase (Hirasawa and Sasaki, 2009).

1.4.2.2 DNA demethylation and re-methylation dynamics in the mammalian germ line

As introduced in Section 1.3.1, following specification from highly methylated epiblast cells, primordial germ cells undergo near complete loss of DNA methylation in two waves. First, a roughly 50% reduction of methylation at promoters, gene bodies, and intergenic regions is observed during PGC migration (Figure 1.4) (Guibert et al., 2012; Seisenberger et al., 2012), although significant methylation is still retained at retrotransposons, imprint control regions and germline gene promoters (Figures 1.3, 1.4) (Seisenberger et al., 2012). Second, following migration into the genital ridge, the bulk of remaining 5mC is lost at most regulatory regions, including germline promoters and ICRs, and repetitive elements, including retrotransposons (Figures 1.3, 1.4, 1.5C) (Guibert et al., 2012; Hackett et al., 2013; Hajkova et al., 2002; Hajkova et al., 2010; Lee et al., 2002; Seisenberger et al., 2012). Similar observations have been seen in gonadal PGCs of other species, with global hypo-methylation, including loss of methylation imprinting, observed in PGCs isolated from human embryos (Gkountela et al., 2015; Guo et al., 2015; Hyldig et al., 2011; Tang et al., 2015).

With respect to DNA demethylation observed in PGCs, both passive (Kagiwada et al., 2013; Seisenberger et al., 2012; Seki et al., 2005) and active models (Hajkova et al., 2010; Kawasaki et al., 2014; Popp et al., 2010) have been proposed. The former mechanism is supported by qPCR and immunostaining analysis of the DNA methyltransferase machinery, which revealed significantly reduced levels of *Dnmt1*, *Dnmt3a*, *Dnmt3b*, and *Uhrf1* in PGCs from E9.5 through to E13.5, along with no detectable expression of *Dnmt3l* and a clear lack of Dnmt1/PCNA co-localisation during S-phase between these stages (Kagiwada et al., 2013; Seisenberger et al., 2012; Seki et al., 2005). Ultimately, this leads to an apparent gradual loss of methylation at imprint control regions (Kagiwada et al., 2013). However, with the exception of *Dnmt3l*, it should be noted that all components of the DNA methyltransferase machinery are clearly detectable (Seisenberger et al., 2012). Moreover, it has

recently been shown that treatment of pregnant mice starting at E9.5 with aphidicolin, which impairs cell cycle progression through S-phase, only mildly affects demethylation of the *H19* imprint control region in the embryonic PGCs, suggesting that this process proceeds via an active mechanism (Kawasaki et al., 2014). In line with this hypothesis, high levels of the Base Excision Repair (BER) components (XRCC1, Ape1, and Parp1) are detected and chromatin bound during gonadal epigenetic reprogramming in PGCs (Hajkova et al., 2010), with significantly increased methylation levels at the H19 imprint control region detected in these cells upon treatment of pregnant mice starting at E9.5 with 3-AB (Kawasaki et al., 2014). With respect to a potential active mechanism driving DNA demethylation in gonadal PGCs, 5mC has been suggested to be modified to thymine by the enzymatic activity of the DNA deaminase Aid, with Aid deficient E13.5 PGCs showing higher levels of DNA methylation than their wild type equivalents (Popp et al., 2010). It should, however, be noted that the observed hyper-methylation is only slight (Popp et al., 2010), and that DNA deaminases (including Aid) generally prefer unmodified cytosines in a single stranded context as their substrate (Nabel et al., 2012).

More recently, and in agreement with the possible involvement of 5hmC in DNA demethylation (see section 1.4.1 for details), 5hmC mapping and gene knockout studies have functionally implicated 5mC oxidation in the DNA demethylation events observed in both migratory and gonadal mouse PGCs (Hackett et al., 2013; Vincent et al., 2013; Yamaguchi et al., 2012; Yamaguchi et al., 2013b). With respect to the former, detectable expression of Tet1 and Tet2 and clearly visible immunofluorescence signal for 5hmC have been observed in migrating PGCs *in vivo* (Hackett et al., 2013; Vincent et al., 2013). However, there is conflicting evidence as to whether these Tet proteins are in fact required for DNA demethylation at this stage. In one study, Surani and colleagues, using a PGC-like cell (PGCLC) system that mimics migratory/pre-reprograming gonadal PGCs *in vitro*, showed by targeted bisulphite sequencing that Tet1/Tet2 depleted PGCLCs had increased methylation at the *Dazl* promoter and repetitive elements compared with their wild type counterparts (Hackett et al., 2013). It should be noted, however, that the bulk of methylation at the particular loci tested is only

lost following gonadal epigenetic reprogramming *in vivo* (i.e. at a later stage of PGC development) (Seisenberger et al., 2012), and thus the functional importance of slightly increased methylation levels in a system which mimics migratory PGCs is unclear. In a second study, Clark and colleagues showed by whole genome bisulphite sequencing (WGBS) that the methylomes of wild type and Tet1/Tet2 depleted induced PGCs (iPGCs), which also mimic migratory/pre-reprograming gonadal PGCs *in vitro*, are largely similar, with only few regions of differential methylation (Vincent et al., 2013). Thus, these results suggest that Tet1/2 oxygenases play potentially only a minor role in the 5mC reduction observed in early mouse germ cells.

With respect to gonadal DNA demethylation in post-migratory mouse PGCs, Surani and colleagues observed globally increased levels of 5hmC in mouse (and human) PGCs by immunofluorescence concurrent with loss of 5mC (Hackett et al., 2013; Tang et al., 2015). Additionally, using a (h)meDIP approach, the authors of this study observed increased levels of 5hmC and decreased levels of 5mC in PGCs during this period at four imprint control regions and at the *Dazl* promoter (Hackett et al., 2013). Again, similar results were also observed in human PGCs during epigenetic reprogramming (Tang et al., 2015). Additionally, studies on mouse germ cells derived from *Tet1* knockout (*Tet1*-KO) and *Tet1*/*Tet2* double knockout (*Tet1*/*Tet2*-DKO) adult mice have further implicated a role for oxidation of 5mC in the DNA demethylation events observed during gonadal germ cell development (Dawlaty et al., 2013; Yamaguchi et al., 2013b). Specifically, Zhang and colleagues showed *Tet1*-KO males have a fertility phenotype consistent with defective genomic imprinting in their offspring, with many embryos derived from *Tet1*-KO males crossed with wild type females dying at mid-gestation (Yamaguchi et al., 2013b). Consistently, de-regulated expression of imprinted genes was observed in these embryos, which was associated with increased methylation at imprint control regions (Yamaguchi et al., 2013b). Additionally, increased methylation at imprint control regions was also observed in the sperm of *Tet1*-KO males and PGCs of E13.5 male and female *Tet1*-KO embryos (Yamaguchi et al., 2013b). Jaenisch and colleagues similarly showed that embryos generated from mating *Tet1*/*Tet2* DKO males or females with *Tet1*/*Tet2* double-heterozygous (Dhet) partners have

increased methylation at many imprint control regions, a phenotype also observed in the sperm of *Tet1*/*Tet2* DKO males (Dawlaty et al., 2013). It should be noted that analysis of imprinted gene expression and imprint control region methylation was not assessed in an allele specific way, and thus it remains unclear if the de-regulated expression or increased methylation is specifically associated with one allele, as would be expected with impaired demethylation of imprint control regions during gonadal epigenetic reprogramming in PGCs.

In addition to the increased levels of methylation observed at imprint control regions in E13.5 *Tet1*-KO PGCs, Zhang and colleagues showed that, in E13.5 *Tet1*-KO female PGCs, increased levels of methylation are observed at the few (i.e. three) germline promoters they tested at high coverage (Yamaguchi et al., 2012). Intriguingly, however, whole genome bisulphite sequencing (WGBS) on E13.5 female *Tet1*-KO PGCs suggested that these cells have globally hypo-methylated DNA (Yamaguchi et al., 2012). It should be noted, however, that with only 2x mean coverage it is difficult to draw meaningful observations or conclusions from this dataset (Yamaguchi et al., 2012). Nonetheless, further RNA-Seq analysis on E13.5 female *Tet1*-KO PGCs, which normally would have just initiated the meiosis programme, revealed significantly reduced expression of meiosis-specific genes (Yamaguchi et al., 2012). Based on these results, and the observation that that *Tet1*-KO female mice showed impaired meiosis and significant reduction in oocyte number starting at E16.5, the authors suggested that Tet1-mediated DNA demethylation of meiosis genes is probably involved in their activation during PGC development (Yamaguchi et al., 2012). However, it should be noted that that some downregulated genes, such as *Stra8*, showed no obvious change in DNA methylation indicating that they are either regulated indirectly by Tet1 or regulated in a DNA-methylation-independent manner (Yamaguchi et al., 2012).

On the basis of the observations described above, two potential models for gonadal DNA demethylation have been put forward. In the first model, comprehensive DNA demethylation in

PGCs, including imprint erasure, entails conversion of 5mC to 5hmC by Tet1 and Tet2, with the rate of progressive 5hmC loss, both globally and at single-copy loci, consistent with a replication-dependent mechanism of DNA demethylation (Figure 1.8A) (Hackett et al., 2013). In the second model, DNA demethylation proceeds globally in the absence of Tet1, but ICRs and germline promoters are demethylated through Tet1-mediated erasure, with 5hmC in PGCs generated by Tet1 diluted in a replication-dependent manner (Yamaguchi et al., 2013b). While both models propose a critical role for Tet1 enzymatic activity in DNA demethylation in gonadal PGCs, neither model has been functionally validated. Rather, the first model infers a role for Tet1 based on 5hmC dynamics during reprogramming in wild type PGCs at a few loci (Hackett et al., 2013), while the second model infers impaired DNA demethylation in the absence of Tet1 based on increased total DNA modification levels (as determined by bisulphite sequencing) in: 1) late stage *Tet1*-KO PGCs and adult germ cells (Yamaguchi et al., 2012; Yamaguchi et al., 2013b); and 2) in the embryonic offspring of *Tet1*-KO males crossed with wild type females (Yamaguchi et al., 2013b).

Following global loss of 5mC, sex-specific DNA methylation patterns are subsequently re-established via the enzymatic activity of Dnmt3a (in males and females) and/or Dnmt3b (males), with the help of their co-factor Dnmt3L (Bourc'his, 2004; Bourc'his et al., 2001; Kaneda et al., 2004; Kato et al., 2007b). In the male germline, all retrotransposon sequences undergo *de novo* DNA methylation during the foetal prospermatogonium stage, concomitant with the *de novo* methylation of paternally methylated ICRs (Bourc'his, 2004; Kaneda et al., 2004; Kato et al., 2007b). In the female germline, the initiation of DNA-methylation at maternally-methylated ICRs and H3K4me3-depleted CpG islands occurs during post-natal oocyte growth (Bourc'his et al., 2001; Kaneda et al., 2004; Smallwood et al., 2011). Additionally, *de novo* DNA methylation is also observed during oogenesis at many repetitive elements, including LINE1 and LTR retrotransposons (Smallwood et al., 2011; Wang et al., 2014). However, in general, it should be noted that methylation levels in the mature-oocyte genome tend to be lower than the levels observed in mature sperm (Wang et al., 2014).

1.5 Research focus: in search of the molecular triggers and mechanisms driving DNA demethylation during gonadal epigenetic reprogramming in mouse PGCs

In view of the observations discussed above, DNA demethylation in gonadal PGCs appears to play a critical role for both licensing primordial germ cells for gametogenesis and for the correct establishment of sex specific methylation patterns at imprint control regions in gametes (Figure 1.5, and as described in Section 1.3.2). However, many questions regarding the molecular mechanisms underlying, and functional consequences of, this DNA demethylation event remain unclear, including: 1) the exact role of Tet1 and 5mC oxidation in global, imprint control region and germline gene promoter DNA demethylation; and 2) the exact role of Tet1 in transcriptional regulation of methylation-sensitive genes during meiosis initiation. With this in mind, I initially set out to functionally validate the role of Tet1 as a putative DNA demethylase during gonadal epigenetic reprogramming in mouse PGCs. In addition, I looked to further characterise the role of Tet1 and 5mC oxidation during transcriptional changes in foetal germ cells. In doing so, I hoped to provide insight into: 1) a (potentially) general mechanism underlying global DNA demethylation in mammals; and 2) a better understanding into the complex interplay between global reprogramming of DNA modifications and transcriptional transitions in an *in vivo* mammalian developmental system.

Chapter 2

Materials and Methods

2.1 Mice

All animal experiments were carried out under a UK Home Office Project Licence in a Home-Office designated facility.

Except for direct comparison with *Tet1*-KO PGCs, wild type PGCs were isolated from embryos produced by crossing mixed background GOF18ΔPE-EGFP (Yoshimizu et al., 1999) transgenic males with either: 1) MF1 females (for AbaSeq, WGBS and immunostaining analysis described in Chapter 3, and the signalling screen and RNA-Seq analysis described in Chapter 5); or 2) C57B6 females (for the organ culture experiments described in Chapter 5). In all cases: 1) the mating is timed in the way that appearance of a vaginal plug at noon is defined as embryonic day (E) 0.5; and 2) the sex of embryos from E12.5 onwards was determined by visual inspection of the gonads.

For study of *Tet1*-KO PGCs, the *Tet1* knockout mouse strain (B6;129S4-Tet1$^{tm1.1Jae}$ /J) (Dawlaty et al., 2011) was purchased from Jackson Laboratory and bred onto the GOF18ΔPE-EGFP (Yoshimizu et al., 1999) transgenic mouse line. Wild type and *Tet1*-knockout (KO) primordial germ cells (PGCs) were isolated from embryos produced from crosses between *Tet1*-heterozygous GOF18ΔPE-EGFP-homozygous females and males. In all cases, the mating is timed in the way that appearance of a vaginal plug at noon is defined as E0.5. For genotyping of embryos produced by crossing *Tet1*-heterozygous GOF18ΔPE-EGFP-homozygous males and females, embryo tails were digested in genotyping buffer (50 mM KCL, 10 mM Tris-HCL, 0.1 mg/ml gelatin, 0.45% Tween 20 and H_2O) and Qiagen protease (Qiagen) at 55°C overnight. Qiagen protease was heat inactivated for 10 min at 95°C. PCR was always carried out twice using two different sets of primers (see Table 2.1) to confirm exon 4 deletion using the RedTaq Ready Mix (Sigma). The sex of the embryos from E12.5 onwards was determined by visual inspection of gonads and additionally confirmed by PCR for *Sry* using the RedTaq Ready Mix (Sigma) and the primers described in Table 2.1.

	Forward primer	Reverse primer
Tet1 (1)	TCAGGGAGCTCATGGAGACTA	TTAAAGCATGGGTGGGAGTC
Tet1 (2)	AACTGATTCCCTTCGTGCAG	TTAAAGCATGGGTGGGAGTC
Sry	TTGTCTAGAGAGCATGGAGGGCCATGTCAA	CCACTCCTCTGTGACACTTTAGCCCTCCGA

Table 2.1 – List of genotyping primers

2.2 Immunostaining

2.2.1 Immunostaining of collected cryosections

Depending on the stage, the following embryonic tissue was first fixed in 2% PFA (in PBS) for 30 min at 4°C: 1) the embryonic hindgut (E8.5); 2) the embryonic trunk (E9.5-E10.5); and 3) the genital ridge (E11.5-E12.5). Following fixation, tissue was washed in PBS three times for 10 min and then incubated in 15% sucrose (in PBS) overnight. After rinsing with 1% BSA (in PBS) the following day, the tissue was embedded in OCT Embedding Matrix (Thermo Scientific Raymond Lamb) and frozen using liquid nitrogen. Samples were subsequently stored at -80°C. A Leica CM 1950 cryostat was used to cut 10 µm sections from the frozen embedded tissue. Sections were settled on poly-lysine slides (Thermo Scientific) and post-fixed with 2% PFA (in PBS) for 3 minutes. After post-fixation, sections were washed three times for 5 min with PBS.

For 5hmC staining, post-fixed sections were first permeabilized for 30 min with 0.5% Triton X-100 (in 1% BSA/PBS), and subsequently treated with RNase A (10 mg/ml; Roche) (in 1% BSA/PBS) for 1 hour at 37°C. Following three 5 min washes with PBS, sections were incubated with 4N HCl for 10-20 min at 37°C to denature genomic DNA, followed by three 10 min washes with PBS. After incubating for 30 min at room temperature in 1% BSA/PBS containing 0.1% Triton X-100, the sections were incubated with one of the combinations of primary antibodies listed in Table 2.2 at 4°C overnight in the same buffer. Sections were subsequently washed three times in 1% BSA/PBS

containing 0.1% Triton X-100 for 5 min and incubated with the corresponding combination of secondary antibodies listed in Table 2.2 in the same buffer for 1 hour in the dark at room temperature. Secondary antibody incubation was followed by three 5 min washes with PBS. DNA was then stained with propidium iodide (PI) (0.25 mg/ml). After a final wash in PBS for 10 min, the sections were mounted with Vectashield (Vector Laboratories).

Primary antibodies	**Secondary antibodies**
Anti-SSEA1 1:2 (gifted by Dr P. Beverly via Dr G. Durcova Hills) anti-5hmC 1:400 (Active motif 39791)	Alexa Fluor 647 Goat Anti-Mouse IgM 1:300 (Invitrogen A21238) Alexa Fluor 488 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11008)
Anti-MVH 1:100 (Abcam 27591) anti-5hmC 1:400 (Active motif 39791)	Alexa Fluor 405 Goat Anti-Mouse IgG 1:300 (Invitrogen A31553) Alexa Fluor 488 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11008)

Table 2.2 – List of primary and secondary antibody combinations used

For all other stainings, post-fixed sections were first incubated for 30 min at room temperature in 1% BSA/PBS containing 0.1% Triton X-100, followed by incubation with one of the combinations of primary antibodies listed in Table 2.3 at 4^{o}C overnight in the same buffer. Sections were subsequently washed three times in 1% BSA/PBS containing 0.1% Triton X-100 for 5 min and incubated with the corresponding combination of secondary antibodies listed in Table 2.2 in the same buffer for 1 hour in the dark at room temperature. Secondary antibody incubation was followed by three 5 min washes with PBS. DNA was then stained with DAPI (0.02 µg/ml, Severn Biotech), and sections were washed in PBS for 10 min and mounted with Vectashield (Vector Laboratories).

Primary antibodies	Secondary antibodies
Anti-Oct3/4 1:200 (BD Biosciences 611202) Anti-P-Akt1 S473 1:50 (Cell Signaling 4060)	Alexa Fluor 568 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11036) Alexa Fluor 488 Goat Anti-Mouse IgG 1:300 (Invitrogen A11001)
Anti-Oct3/4 1:200 (BD Biosciences 611202) Anti-P-p44/42 MAPK (Erk1/2) 1:200 (Cell Signaling 4370)	Alexa Fluor 568 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11036) Alexa Fluor 488 Goat Anti-Mouse IgG 1:300 (Invitrogen A11001)
Anti-Oct3/4 1:200 (BD Biosciences 611202) Anti-P-Smad1/5 1:50 (Cell Signaling 9516)	Alexa Fluor 568 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11036) Alexa Fluor 488 Goat Anti-Mouse IgG 1:300 (Invitrogen A11001)
Anti-Oct3/4 1:200 (BD Biosciences 611202) Anti-H3S10P 1:200 (Abcam 5176)	Alexa Fluor 568 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11036) Alexa Fluor 488 Goat Anti-Mouse IgG 1:300 (Invitrogen A11001)
Anti-Oct3/4 1:200 (BD Biosciences 611202) Anti-P-4EBP 1:200 (Cell Signaling 2855)	Alexa Fluor 568 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11036) Alexa Fluor 488 Goat Anti-Mouse IgG 1:300 (Invitrogen A11001)
Anti-Oct3/4 1:200 (BD Biosciences 611202) Anti-Foxo1 1:100 (Cell Signaling 2880)	Alexa Fluor 568 Goat Anti-Rabbit IgG 1:300 (Invitrogen A11036) Alexa Fluor 488 Goat Anti-Mouse IgG 1:300 (Invitrogen A11001)
Anti-GFP 1:1000 (Abcam 5450) Anti-P-4EBP 1:200 (Cell Signaling 2855)	Alexa Fluor 568 Donkey Anti-Rabbit IgG 1:300 (Invitrogen A10042) Alexa Fluor 488 Donkey Anti-Goat IgG 1:300 (Invitrogen A11055)
Anti-GFP 1:1000 (Abcam 5450) Anti-Dazl 1:300 (Abcam 34139)	Alexa Fluor 568 Donkey Anti-Rabbit IgG 1:300 (Invitrogen A10042) Alexa Fluor 488 Donkey Anti-Goat IgG 1:300 (Invitrogen A11055)

Table 2.3 – List of primary and secondary antibody combinations used

2.2.2 Immunostaining of dissociated cells

Depending on the stage, the following embryonic tissue was first digested at 37ºC for 3 min using 0.05% Trypsin-EDTA (1x) (Gibco) or TrypLE Express (Thermo): 1) the embryonic hindgut (E8.5); 2) the embryonic trunk (E9.5-E10.5); and 3) the genital ridge (E11.5-E12.5). Enzymatic digestion was followed by neutralization with DMEM/F-12 (Gibco) containing 15% foetal bovine serum (Gibco) and pipetting up and down 10-20 times, followed by centrifugation (3 min, 2,000 rpm) to collect cells. Cells were subsequently resuspended in DMEM/F-12 supplemented with hyaluronidase (300 µg/ml; Sigma) by pipetting up and down 20-30 times to generate single cell suspensions. Following centrifugation (3 min, 2,000 rpm) to collect cells, the cell pellet was resuspended in DMEM/F-12 (Gibco). The single cell suspension was plated on a poly-lysine slide and allowed to settle and attached for 15 min, at which point media was gently removed with a pipette. Cells were subsequently

fixed with 2% PFA (in PBS) for 10 min and washed three times with PBS. Fixed cells were then incubated for 30 min at room temperature with 1% BSA/PBS containing 0.1% Triton X-100, followed by incubation with primary antibodies (anti-active β-catenin (1:500, Millipore 05-665) and anti-GFP (1:1000, Abcam 5450)) at 4^{o}C overnight in the same buffer. Cells were subsequently washed three times in 1% BSA/PBS containing 0.1% Triton X-100 for 5 min and incubated with secondary antibodies (Alexa Fluor 568 donkey anti-mouse IgG (1:300, Invitrogen A10037) and Alexa Fluor 488 donkey anti-goat IgG 1:300 (1:300, Invitrogen A11055)) for 1 hour in the dark at room temperature in the same buffer. Secondary antibody incubation was followed by three 5 min washes with PBS. DNA was then stained with DAPI (0.02 µg/ml, Severn Biotech), and sections were washed in PBS for 10 min and mounted with Vectashield (Vector Laboratories).

2.3 PGC isolation by flow cytometry

Depending on the stage, the following embryonic tissue was first digested at 37ºC for 3 min using 0.05% Trypsin-EDTA (1x) (Gibco) or TrypLE Express (Thermo): 1) the embryonic hindgut (E8.5); 2) the embryonic trunk (E9.5-E10.5); and 3) the genital ridge (E11.5-E14.5). Enzymatic digestion was followed by neutralization with DMEM/F-12 (Gibco) containing 15% foetal bovine serum (Gibco) and pipetting up and down 10-20 times, followed by centrifugation (3 min, 2,000 rpm) to collect cells. Cells were subsequently resuspended in DMEM/F-12 supplemented with hyaluronidase (300 µg/ml; Sigma) by pipetting up and down 20-30 times to generate single cell suspensions. Following centrifugation (3 min, 2,000 rpm) to collect cells, the cell pellet was resuspended in ice-cold PBS supplemented with poly-vinyl alcohol (10 µg/ml) and EGTA (0.4 mg/ml, Sigma). GFP positive cells were finally isolated using an Aria IIu (BD Bioscience) or Aria III (BD Bioscience) flow cytometer (done in collaboration with Dr James Elliot, MRC Clinical Sciences Centre FACS facility), and sorted into ice cold PBS supplemented with poly-vinyl alcohol (10 µg/ml) and EGTA (0.4 mg/ml, Sigma). GFP positive cells were isolated with >95% purity.

2.4 Next generation sequencing library preparation

2.4.1 RNA-Seq library prerparation

For study of *Tet1*-KO PGCs, total RNA from sorted PGCs isolated from individual *Tet1*-KO or wild type embryos was isolated using ZR-Duet DNA-RNA MiniPrep kit (Zymo), and RNA from between two to six embryos (equivalent to 1,000 to 8,000 cells) of the same genotype, stage and sex was pooled and concentrated to 6 µL final volume using the RNA Clean and Concentrator 5 kit (Zymo). For study of wild type PGCs isolated from embryos produced by crossing MF1 females with GOF18ΔPE-EGFP males, total RNA from 600-1,000 sorted PGCs from pooled litters of embryos was isolated using the Nucleospin RNA XS kit (Macherey-Nagel). cDNA synthesis and amplification (15 cycles) was performed with the SMARTer Ultra Low Input RNA kit (Clontech) (Ramsköld et al., 2012), following the manufacturer's instructions. The amplified cDNA was fragmented by a Covaris S2 sonicator (Covaris), following the manufacturer's instructions, and sheared cDNA was converted to sequencing libraries using the NEBNext Ultra DNA Library Prep kit (NEB), following the manufacturer's instructions and using 13 cycles of amplification and NEBNext Multiplex Oligos for Illumina sequencing (NEB). The libraries were purified by AMPure XP beads (Beckman-Coulter), and bar-coded libraries were pooled and sequenced on the Illumina HiSeq 2000 instrument.

2.4.2 AbaSeq library preparation

Total DNA was isolated from 10,000 sorted PGCs using the QIAamp DNA Micro Kit (Qiagen) using 0.8 µg of carrier RNA and following the "Isolating DNA from small samples of blood" protocol. AbaSeq libraries for 5hmC profiling were constructed as previously described (Sun et al., 2013). In brief, genomic DNA was glucosylated with 400 nmol UDP-glucose and 10 units of T4-β-glucosyltransferase (NEB) in NEB buffer 4 overnight at 37°C and was purified afterward by phenol-

chloroform extraction and ethanol precipitation. The glucosylated genomic DNA was then digested by four units of AbaSI enzyme (NEB) in NEB buffer 4 at room temperature (25°C) for 1 hr and the enzyme was inactivated at 65°C for 20 minute. Biotinylated P1 adapters were ligated onto the AbaSI digested DNA by incubating the digested DNA with 50 pmol of P1 and 2 µl of T4 DNA quick ligase (NEB) in the 1X quick ligation buffer in a total volume of 42 µl at room temperature for 30 min. Ligated DNA was then purified by AMPure XP beads (Beckman-Coulter) and was fragmented using a Covaris S2 sonicator (Covaris), following the manufacturer's instructions. Sheared P1-ligated DNA was then captured by mixing with 12.5 µl of Dynabeads MyOne Streptavidin C1 beads (Life Technologies) according to the manufacturer's specifications. End repair and dA-tailing were carried out on the beads by using the NEBNext End Repair Module (NEB) and the NEBNext dA-tailing Module (NEB) at 20°C and 37 °C respectively for 30 min. The beads were washed twice with a buffer containing 5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, and 1 M NaCl between the two reactions and after dA-tailing. P2 adapters were ligated to the random sheared ends of the dA-tailed DNA using NEB T4 DNA quick ligase at room temperature for 30 min. After washing the DNA-bound beads as described above, the entire DNA was amplified using the Phusion DNA polymerase (NEB) with the addition of 300 nM forward primer (PCR_I) and 300 nM reverse primers (PCR_IIpe) for 16 cycles. The beads were then removed from the PCR and the library was purified using AMPure XP beads (Beckman-Coulter). Libraries were sequenced on the Illumina HiSeq 2000 instrument. Of note, AbaSeq library preparation from DNA I isolated was carried out by Dr Jolyon Terragni (New England Biolabs (NEB), MA, USA) under the supervision of Dr Sriharsa Pradhan (NEB, MA, USA) and Dr Yu Zheng (NEB, MA, USA).

2.4.3 Whole genome bisulphite sequencing (WGBS) library preparation

Total DNA was isolated from 10,000 sorted PGCs using the QIAamp DNA Micro Kit (Qiagen) using 0.8 µg of carrier RNA and following the "Isolating DNA from small samples of blood" protocol. For

the second biological replicates, unmethylated λ phage DNA (Promega) was spiked in following DNA isolation to assess bisulphite conversion rate. DNA was fragmented using a Covaris S2 sonicator (Covaris), as per manufacturer's instructions. Libraries were made following the NEBNext Library Prep protocol, with the following modifications: 1) methylated adaptors were diluted 1:10 in nuclease free H_2O to a final concentration of 1.5 μM; 2) following adaptor ligation, bisulphite conversion was carried out using the Imprint Modification Kit (Sigma), and following the two step protocol according to the manufacturer's specifications; and 3) PCR enrichment was carried out for 16 cycles using the NEXTflex Bisulphite-Seq Kit for Illumina Sequencing (Bioo Scientific) master mix and the NEBNext Library Prep universal and index primers (NEB). The libraries were purified by AMPure XP beads (Beckman-Coulter). Libraries were sequenced on the Illumina HiSeq 2000 instrument.

2.4.4 Reduced representation bisulphite sequencing (RRBS) library preparation

Total DNA from FACS-sorted PGCs isolated from individual *Tet1*-KO or wild type embryos was isolated using ZR-Duet DNA-RNA MiniPrep kit (Zymo), and DNA from between two to six embryos (equivalent to 1,000 to 8,000 cells) of the same genotype, stage and sex was pooled and concentrated to 26 μL final volume using the Savant SpeedVac Concentrator (Thermo), following the manufacturer's instructions. Genomic DNA was digested by 20 units of MspI enzyme (NEB) in NEB buffer 2 at 37°C for 3 hrs, and digested DNA was purified using AMPure XP beads (Beckman-Coulter). Libraries were made following the NEBNext Ultra DNA Library Prep protocol, with the following modifications: 1) methylated adaptors were diluted 1:10 in nuclease free H_2O to a final concentration of 1.5 μM; 2) following adaptor ligation, bisulphite conversion was carried out using the Imprint Modification Kit (Sigma), and following the two step protocol according to the manufacturer's specifications; and 3) PCR enrichment was carried out for 18 cycles using the KAPA Uracil$^+$ DNA polymerase master mix (KAPA Biosystems) and the NEBNext Library Prep universal

and index primers (NEB). The libraries were purified by AMPure XP beads (Beckman-Coulter). Pooled libraries were sequenced on the Illumina HiSeq 2000 instrument.

2.5 Bioinformatics

2.5.1 Published datasets used

In addition to the original next generation sequencing (NGS) datasets generated (see Section 2.4), the following published NGS datasets were re-analysed:

Library strategy	Cell type	Genotype	Accession	Original reference
Tet1 ChIP-Seq	Mouse ESC	Wild type	GSE24843	ref. (Williams et al., 2011)
Sin3a ChIP-Seq	Mouse ESC	Wild type	GSE24843	ref. (Williams et al., 2011)
OGT ChIP-Seq	Mouse ESC	Wild type	GSE39154	ref. (Vella et al., 2013)
Ezh2 ChIP-Seq	Mouse ESC	Wild type	GSE49435	ref. (Kaneko et al., 2013)
WGBS	Mouse ESC	Wild type	GSE48519	ref.(Hon et al., 2014)
WGBS	Mouse ESC	*Tet1*-KO	GSE48519	ref.(Hon et al., 2014)
TAB-Seq	Mouse ESC	Wild type	GSE48519	ref.(Hon et al., 2014)
MAB-Seq	Mouse ESC	Wild type	GSE55660	ref.(Neri et al., 2015)
RNA-Seq	Mouse ESC	*Tet1*-KO	GSE48519	ref.(Hon et al., 2014)
RNA-Seq	Mouse ESC	Wild type	GSE48519	ref.(Hon et al., 2014)
RNA-Seq	Mouse ESC	*Tdg*-KO	GSE41545	ref. (Song et al., 2013a)
RNA-Seq	Mouse ESC	Wild type	GSE41545	ref. (Song et al., 2013a)
H3K4me3 ChIP-Seq	Day 6 PGCLC	Wild type	GSE60204	ref.(Kurimoto et al., 2015)
H3K27Ac ChIP-Seq	Day 6 PGCLC	Wild type	GSE60204	ref.(Kurimoto et al., 2015)
H3K27me3 ChIP-Seq	Day 6 PGCLC	Wild type	GSE60204	ref.(Kurimoto et al., 2015)
RNA-Seq	E13.5 female PGC	*Tet1*-KO	GSE41908	ref.(Yamaguchi et al., 2012)
RNA-Seq	E13.5 female PGC	Wild type	GSE41908	ref.(Yamaguchi et al., 2012)
hMeDIP	E10.5 PGC	Wild type	SRA060914	ref.(Hackett et al., 2013)
MeDIP	E10.5 PGC	Wild type	SRA060914	ref.(Hackett et al., 2013)
AbaSeq	Mouse ESC	Wild type	GSE42898	ref.(Sun et al., 2013)

Table 2.4 – Summary of published datasets re-analysed

2.5.2 RNA-Seq alignment and downstream analysis

For study of *Tet1*-KO and *Tet1*-WT PGCs (Chapter 4), Illumina and Smart-seq adapters from the sequencing reads were first trimmed using Trimmomatic. Trimmomatic produces three fastq output

files, one with paired end reads (PE) and two (F and R) with reads whose mate did not pass trimming criteria; PE, F and R fastq files were aligned separately and BAM files were merged at the end. For other RNA-Seq libraries, fastq files generated from output of next generation sequencing were used directly for alignment. RNA-Seq reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (version 0.12.8) (Langmead et al., 2009) and Tophat (version 2.0.2) (Trapnell et al., 2009) with options `-N 2 --b2-very-sensitive --b2-L 25` and `--library-type` set according to library preparation method. Annotations from Ensembl Gene version 67 were used as gene model with Tophat. Read counts per annotated gene were computed using HTSeq (version 0.5.3p9) (Anders et al., 2015) and expression level of each gene was quantified by computing the number of fragments detected per kilobase per million of reads (FPKM) using custom R script. Differential expression analysis was performed using DESeq2 (version 1.6.3) (Love et al., 2014). For all analyses, genes were considered differentially expressed if their adjusted p-value < 0.05, as computed by DESeq2 software. Gene ontology analysis of differentially expressed genes was performed using DAVID (http://david.abcc.ncifcrf.gov) for unranked gene lists or using Gene Set Enrichment Analysis (GSEA, version 2.2.0) (Subramanian et al., 2005) for ranked genes lists. For GSEA, the gene list was ranked based on both the adjusted p-value (i.e. '1 - (adjusted p-value)') and fold-change (i.e. up-regulated genes were multiplied by a factor of '1'; down-regulated genes were multiplied by a factor of '-1') computed from DESeq2 differential expression analysis.

Of note, depending on the dataset, initial alignment of RNA-Seq reads was done by either Gopuraja Dharmalingam (MRC Clinical Sciences Centre, UK) or me. All further downstream analysis was carried out by me.

2.5.3 Whole genome bisulphite sequencing (WGBS) and Tet-assisted bisulphite sequencing (TAB-Seq) alignment and downstream analysis

For all whole genome bisulphite sequencing (WGBS) and Tet-assisted bisulphite sequencing (TAB-Seq) datasets, raw reads were first trimmed using Trim Galore (version 0.3.1, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters for single end reads and the `--paired --trim1` options for paired end reads. Alignments were carried out to the mouse genome (mm9, NCBI build 37) with Bismark (version 0.13.0) (Krueger and Andrews, 2011) with the `-n 1` parameter; where appropriate, the λ phage genome was added as an extra chromosome. Aligned reads were deduplicated with deduplicate_bismark selecting a random alignment for positions that were covered more than once. Where appropriate, the bisulphite conversion rate was computed using reads aligned to the λ phage genome and using the to-mr script (parameters: `-m bismark`) and bsrate script (paramters: `-N`) of Methpipe (version 3.3.1) (Song et al., 2013b). CpG methylation calls were extracted from the deduplicated mapping output using the Bismark methylation extractor (version 0.13.0) with the `--no_overlap --comprehensive --merge_non_CpG` parameters for paired end mode and `--comprehensive --merge_non_CpG` parameters for single end mode. The number of methylated and unmethylated cytosines in a CpG context was extracted using bismark2bedGraph (default parameters) and coverage2cytosine (parameters: `--zero_based`). Symmetric CpGs were merged with custom R script. For all downstream analysis, only symmetric CpGs with minimum 8x (WGBS) or 16x (TAB-Seq) coverage were used. Unless stated otherwise, all WGBS and TAB-Seq analysis was carried out on data from merged biological replicates for ease of data visualisation.

The mapBed function of BEDtools (version 2.24.0) (Quinlan and Hall, 2010) was used to compute the mean total DNA modification level (WGBS) or 5hmC level (TAB-Seq) for the following genomic features: 1) all 2kb windows (containing a minimum 4 symmetric CpGs); 2) gene promoters (defined as Ensembl 67 gene start sites -1kB/+500bp); 3) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites); 4) putative active enhancers in day 6 PGCLCs (Kurimoto et al., 2015) (as defined in Section 2.5.6); 5) imprint control regions (the coordinates of which for the mm9 genome were kindly provided by Dr Kirsten McEwen (MRC Clinical Science Centre, UK); 6)

intergenic regions (defined as all regions not overlapping any of the aforementioned regions or any element identified by the RepeatMasker track (Fujita et al., 2011) of mm9 (UCSC Genome Browser)); and 7) metagene plot bins (for how these are defined, please refer to the description at the end of this of this paragraph). In some instances, promoters were divided into high CpG (HCP), intermediate CpG (ICP) and low CpG (LCP) promoters, as defined using the same parameters as previously published (Borgel et al., 2010). Briefly, LCPs contain no 500-bp window with a CpG ratio > 0.45; HCPs contain at least one 500-bp window with a CpG ratio >0.65 and GC content > 55%; ICPs do not meet the previous criteria. For metagene plots, a genomic feature was divided into equally sized bins using BEDtools makewindows. The genomic features used for this analysis include: 1) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites) +/- 0.5 times the length of the gene body (100 bins); 2) methylation canyons and cUMRs (as defined below) +/- 1 times the length of the methylation canyon/cUMR (90 bins); 3) putative active enhancers in day 6 PGCLCs (Kurimoto et al., 2015) (as defined in Section 2.5.6) +/- 1 times the length of the putative active enhancer (90 bins); and 4) transcription start sites (defined as Ensembl 67 gene start sites) +/- 2 kB (20 bins). The total DNA modification level (WGBS) and 5hmC level (TAB-Seq) for all regions was expressed as the mean of individual CpG sites.

Methylation canyons and comparatively unmethylated regions (cUMRs) in E10.5 PGCs were identified first using (in order) the to-mr (parameters: `-m bismark`), methcounts (parameters: `-n`), symmetric-cpgs (default parameters), and hmr scripts (default parameters) of Methpipe (version 3.3.1) (Song et al., 2013b) to identify hypo-methylated regions, and then using the mapBed function from BEDtools (version 2.24.0) (Quinlan and Hall, 2010) to compute the mean total DNA modification level per identified hypo-methylated region. Methylation canyons were defined as hypo-methylated regions greater than 3.5 kB with less than 5% average total DNA modification in E10.5 PGCs; cUMRs were defined as hypo-methylated regions between 1.0 and 3.5 kB with less than 5% average total DNA modification in E10.5 PGCs.

RnBeads (version 1.0.0) (Assenov et al., 2014) and RnBeads.mm9 (version 0.99.0) (Assenov et al., 2014) were used to identify promoters (defined as Ensembl 67 gene start sites -1kB/+500bp) differentially methylated between *Tet1*-KO and wild type mESCs with the following options:

```
rnb.options(assembly="mm9", import.bed.style="bismarkCov",
region.aggregation = "mean",
filtering.snp="yes",filtering.missing.value.quantile=0.95,
filtering.coverage.threshold=8,
filtering.high.coverage.outliers=TRUE,
filtering.low.coverage.masking=TRUE,
normalization.method="none",
normalization.background.method="none",
filtering.context.removal=c("CC","CAG","CAH","CTG","CTH",
"Other"), filtering.sex.chromosomes.removal=FALSE, qc=FALSE,
exploratory=FALSE, differential=TRUE,
differential.adjustment.celltype=FALSE)
```

The following was extracted from the output of RnBeads: 1) the difference in methylation means between the two genotypes for each commonly covered promoter; and 2) the p-value representing the significance of the difference in methylation means between the two genotypes for each commonly covered promoter. Differentially methylated regions were identified as regions with a p-value < 0.05 (as computed by RnBeads programme) and a difference in methylation means between the two genotypes greater than 10% (as computed by RnBeads programme).

For assessing DNA modification levels at specific repetitive elements, Bismark (version 0.14.4) (Krueger and Andrews, 2011) was used to map all reads from each data set against consensus sequences constructed from Repbase (Jurka et al., 2005) (and kindly provided by Dr Felix Krueger (Babraham Institute, UK)) with the `-n 1` parameter set. CpG methylation calls were extracted from the mapping output using the Bismark methylation extractor (version 0.14.4) with the `--no_overlap --comprehensive --merge_non_CpG` parameters set for paired end mode. The number of methylated and unmethylated cytosines in a CpG context were extract using

bismark2bedGraph (default parameters) and coverage2cytosine (with `--zero_based` parameter). The methylation level was expressed as the mean of individual CpG sites.

2.5.4 AbaSeq alignment and downstream analysis

For the uniquely mappable part of the genome, AbaSeq reads were processed as previously described (Sun et al., 2013). In brief, raw sequencing reads were trimmed for adaptor sequences and low quality bases using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The trimmed reads were mapped to the mouse genome (mm9, NCBI build 37) using Bowtie (version 0.12.8) (Langmead et al., 2009) with parameters `-n 1 -l 25 --best --strata -m 1`. Calling of 5hmC was based on the recognition sequence and cleavage pattern of the AbaSI enzyme (5′-$CN_{11–13}$↓$N_{9–10}$G-3′/3′-$GN_{9–10}$↓$N_{11–13}$C-5′) using custom Perl scripts. For assessing relative enrichment of 5hmC at repetitive elements and non-repetitive elements, AbaSeq alignments were divided into two groups: unique (single best alignment) and ambiguous (map to multiple locations with equal alignment score). Both groups were then mapped to the repetitive elements defined by the RepeatMasker track (Fujita et al., 2011) of mm9 (UCSC Genome Browser) separately.

For quantification of relative 5hmC levels at symmetric CpGs in the uniquely mappable part of the genome, the number of counts per symmetric CpG for a given sample were normalised to the combined number of uniquely mapped and ambiguously mapped reads for a given library, and then further multiplied by a stage-specific normalisation factor based on the mean 5hmC level for each stage as computed by LC/MS (E10.5 = 1.0; E11.5 = 1.13; E12.5F = 0.76; E12.5M = 1.0). All symmetric CpGs falling within genomic intervals blacklisted by the mouse (mm9) ENCODE project (ENCODE Project Consortium, 2012) were excluded from all further downstream analysis. Unless stated otherwise, all AbaSeq analysis was carried out on data from merged biological replicates for ease of data visualisation.

The mapBed function of BEDtools (version 2.24.0) (Quinlan and Hall, 2010) was used to compute the mean 5hmC level for the following genomic features: 1) all 2kb windows (containing a minimum 4 symmetric CpGs); 2) gene promoters (defined as Ensembl 67 gene start sites -1kB/+500bp); 3) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites); 4) putative active enhancers in day 6 PGCLCs (Kurimoto et al., 2015) (as defined in Section 2.5.6); 5) imprint control regions (the coordinates of which for the mm9 genome were kindly provided by Dr Kirsten McEwen (MRC Clinical Science Centre, UK); 6) intergenic regions (defined as all regions not overlapping any of the aforementioned regions or any element identified by the RepeatMasker track (Fujita et al., 2011) of mm9 (UCSC Genome Browser)); and 7) metagene plot bins (for how these are defined, please refer to the description at the end of this of this paragraph). In some instances, promoters were divided into high CpG (HCP), intermediate CpG (ICP) and low CpG (LCP) promoters, as defined using the same parameters as previously published (Borgel et al., 2010). Briefly, LCPs contain no 500-bp window with a CpG ratio > 0.45; HCPs contain at least one 500-bp window with a CpG ratio >0.65 and GC content > 55%; ICPs do not meet the previous criteria. For metagene plots, a genomic feature was divided into equally sized bins using BEDtools makewindows. The genomic features used for this analysis include: 1) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites) +/- 0.5 times the length of the gene body (100 bins); 2) methylation canyons and cUMRs (as defined in Section 2.5.3) +/- 1 times the length of the methylation canyon/cUMR (90 bins); 3) putative active enhancers in day 6 PGCLCs (Kurimoto et al., 2015) (as defined in Section 2.5.6) +/- 1 times the length of the putative active enhancer (90 bins); and 4) transcription start sites (defined as Ensembl 67 gene start sites) +/- 2 kB (20 bins). Unless otherwise indicated, the 5hmC level for all regions was expressed as the mean of individual CpG sites.

To identify 5hmC enriched or depleted regions in E10.5 PGCs, I first divided the genome into 2kb windows (minimum 4 symmetric CpGs), and computed the mean 5hmC level for each window using BEDtools (version 2.24.0) (Quinlan and Hall, 2010). I next computed the mean of the mean 5hmC level for all 2kB windows using base R giving value λ. To determine the significance of 5hmC enrichment in each 2kB window, I calculated the p-value based on the upper-tail (to determine 5hmC enriched regions) or lower-tail (to determine 5hmC depleted regions) of the Poisson probability using base R:

$$f(x;\lambda) = \Pr(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

x = observed 5hmCG mean value for each 2kb window

λ = mean of observed 5hmCG mean values for all 2kb windows

To adjust for multiple testing, I applied Benjamini-Hochberg correction (using R) to both the upper-tail and lower-tail p-values, giving a final adjusted p-value for each 2kb window. Windows with adjusted upper-tail p-value < 0.05 were considered relatively enriched for 5hmC while windows with adjusted lower-tail p-value < 0.05 were considered relatively depleted for 5hmC.

For assessing relative enrichment of 5hmC at specific repetitive elements, Bowtie (Langmead et al., 2009) was used to map all reads from each data set against consensus sequences constructed from Repbase (Jurka et al., 2005) (and kindly provided by Dr Felix Krueger (Babraham Institute, UK)) with parameters `-n 1 -M 1 --strata –best`. The number of reads mapped to each sequence within a given sample was first normalised to the library size of that particular sample, and then normalised to the mean proportion of reads mapped to a given sequence in E10.5 PGCs.

Of note, the initial AbaSeq alignment step was carried out by Dr Zhiyi Sun (New England Biolabs, MA, USA). All other data manipulation and downstream analysis was carried out by me.

2.5.5 Reduced representation bisulphite sequencing (RRBS) alignment and downstream analysis

Although reduced representation bisulphite sequencing (RRBS) libraries were sequenced as paired end libraries, FastQC analysis revealed poor sequencing quality of read 2 compared with read 1. All RRBS libraries generated were thus treated as single end libraries with only read 1 used. Raw read 1 reads were first trimmed using Trim Galore (version 0.3.1, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with `--rrbs` parameter. Alignments were carried out to the mouse genome (mm9, NCBI build 37) with Bismark (version 0.13.0) (Krueger and Andrews, 2011) with the `-n 1` parameter. CpG methylation calls were extracted from the mapping output using the Bismark methylation extractor (version 0.13.0) with the `-s --comprehensive --merge_non_CpG` parameters. The number of methylated and unmethylated cytosines in a CpG context was extracted using bismark2bedGraph (default parameters).

RnBeads (version 1.0.0) (Assenov et al., 2014) and RnBeads.mm9 (version 0.99.0) (Assenov et al., 2014) were used to identify differentially methylated regions between two test groups for the following genomic features: 1) all 2kb windows (containing a minimum 4 symmetric CpGs); 2) gene promoters (defined as Ensembl 67 gene start sites -1kB/+500bp); and 3) imprint control regions (the coordinates of which for the mm9 genome were kindly provided by Dr Kirsten McEwen (MRC Clinical Science Centre, UK). The options used for this analysis are listed below:

```
rnb.options(assembly="mm9", import.bed.style="bismarkCov",
region.aggregation = "mean",
filtering.snp="yes",filtering.missing.value.quantile=0.95,
```

```
filtering.coverage.threshold=8,
filtering.high.coverage.outliers=TRUE,
filtering.low.coverage.masking=TRUE,
normalization.method="none",
normalization.background.method="none",
filtering.context.removal=c("CC","CAG","CAH","CTG","CTH",
"Other"), filtering.sex.chromosomes.removal=FALSE, qc=FALSE,
exploratory=FALSE, differential=TRUE,
differential.adjustment.celltype=FALSE)
```

The following was extracted from the output of RnBeads: 1) the mean methylation level for each group (i.e. stage, sex and/or genotype) for each commonly covered test region; 2) the difference in methylation means between two groups for each commonly covered test region; and 3) the p-value representing the significance of the difference in methylation means between two groups for each commonly covered test region. Differentially methylated regions were identified as regions with a p-value < 0.05 (as computed by RnBeads programme) and a difference in methylation means between two groups greater than 10% (as computed by RnBeads programme).

2.5.6 ChIP-Seq, hMeDIP-Seq and MeDIP-Seq alignment and downstream analysis

ChIP-Seq, hMeDIP-Seq and MeDIP-Seq reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (version 0.12.8 or version 1.0.0) (Langmead et al., 2009) with parameters `-n 2 -l 25 -m 1` and `-C` where appropriate. Unless stated otherwise, all ChIP-Seq, hMeDIP-Seq and MeDIP-Seq analyses was carried out on data from merged biological replicates. For transcription regulator and chromatin modifiers (i.e. Tet1, Sin3a, Ezh2, OGT), binding sites were called using MACS (version 2.1.0) (Zhang et al., 2008) callpeak with parameter `-g mm -q 0.01`. Of note, for Tet1 ChIP-Seq in mESCs (Williams et al., 2011) only, peaks were called for the two biological replicates separately, and only overlapping peaks (as determined by BEDtools intersect (Quinlan and Hall, 2010)) were used for downstream analysis. BEDtools intersect (Quinlan and Hall, 2010) was used to identify Tet1 peaks overlapping Sin3a, OGT, or Ezh2 peaks in mESCs. Genes associated with

Tet1 binding sites were determined using Genomic Regions Enrichment of Annotations Tool (GREAT, version 3.0.0) (McLean et al., 2010) using basal (-5kB/+1kB) plus extension (50kb) settings.

To identify putative active enhancers in day 6 PGCLCs (Kurimoto et al., 2015), I first generated an 8-state chromatin model using ChromHMM (Ernst and Kellis, 2012). Putative active enhancers were defined as all regions contained within state 7 ($H3K27Ac^{+}/H3K4me3^{-}/H3K27me3^{-}$; see Figure 2.1 for details) not overlapping any potential promoter regions (Ensembl 67 gene start sites -1kB/+500bp).

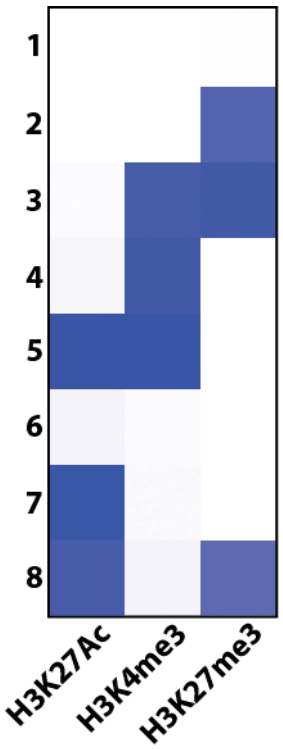


Figure 2.1 – ChromHMM 8-state chromatin model – day 6 PGCLCs (Kurimoto et al., 2015)

For computing transcription regulator or chromatin modifier binding levels from ChIP-Seq datasets around transcription start sites (TSS), the genomic interval around the TSS +/- 2kB was divided into 20 equally sized bins using BEDtools makewindows. I then used BEDtools multicov to compute the number of test and control reads overlapping each bin. The total number of test and control reads per

bin for each sample were normalised to the appropriate library size, and fold enrichment for each bin was determined by dividing the number of normalised ChIP-Seq test sample reads by the number of normalised ChIP-Seq control sample reads.

For hMeDIP and MeDIP, I used BEDtools multicov to identify the number of test reads overlapping each promoter region (defined as Ensembl 67 gene start sites -1kB/+500bp). In some instances, promoters were divided into high CpG (HCP), intermediate CpG (ICP) and low CpG (LCP) promoters, as defined using the same parameters as previously published (Borgel et al., 2010). Briefly, LCPs contain no 500-bp window with a CpG ratio > 0.45; HCPs contain at least one 500-bp window with a CpG ratio >0.65 and GC content > 55%; ICPs do not meet the previous criteria.

2.5.7 Methylation-assisted bisulphite sequencing (MAB-Seq) downstream analysis

For MAB-Seq of mESCs (Neri et al., 2015), the list of previously identified 5fC/5caC enriched CpGs was downloaded from GSE55660.

2.6 Mass spectrometry

Genomic DNA from between 100 and 10,000 FACS-sorted PGCs was extracted using ZR-Duet DNA/RNA Miniprep kit (Zymo Reasearch) following manufacturer instructions and eluted in LC/MS grade water. DNA was digested to nucleosides using a digestion enzyme mix provided by NEB. A dilution-series made with known amounts of synthetic nucleosides and the digested DNA were spiked with a similar amount of isotope-labelled nucleosides (provided by Dr T. Carell (LMU, Germany)) and separated on an Agilent RRHD Eclipse Plus C18 2.1 × 100 mm 1.8u column by using the

UHPLC 1290 system (Agilent) and an Agilent 6490 triple quadrupole mass spectrometer. To calculate the quantity of individual nucleosides, standard curves representing the ratio of unlabelled over isotope-labelled nucleosides were generated and used to convert the peak-area values to corresponding quantity. Threshold for quantification is a signal-to-noise (calculated with a peak-to-peak method) above 10. Of note, mass spectrometry on DNA I isolated was carried out by Dr Rachel Amouroux (MRC Clinical Sciences Centre, UK).

2.7 Organ culture

For 24 hrs organ culture experiment, E10.5 embryonic trunks containing nascent urogenital ridges were cultured *ex vivo*. 2% agar dissolved in DMEM/F12 (Gibco) was poured into the centre well of an organ culture dish (BD Falcon) with a thin glass capillary in the centre to generate a small groove. Trunks were placed into the groove and covered with medium containing DMEM/F12 (Gibco), 15% FBS (Gibco), NEAA (Gibco), glutamine (Thermo), sodium pyruvate (Sigma), Penicillin/Streptomycin (100U/ml) (Thermo), and either 1 μM rapamycin (in EtOH) or equal volume of EtOH. Media was replaced after 6 hrs and again after 18 hrs, 24 hrs and 40 hrs.

For 48 hrs organ culture experiment, E10.5 nascent urogenital ridges were cultured *ex vivo*. 2% agar dissolved in DMEM/F12 (Gibco) was poured into the centre well of an organ culture dish (BD Falcon) with a thin glass capillary in the centre to generate a small groove. Trunks were placed into the groove and covered with medium containing DMEM/F12 (Gibco), 15% FBS (Gibco), NEAA (Gibco), glutamine (Thermo), sodium pyruvate (Sigma), Penicillin/Streptomycin (100U/ml) (Thermo), and either 200 nM rapamycin (in EtOH) or an equal volume of EtOH. Media was replaced after 6 hrs and again after 18 hrs.

Chapter 3

Results

On the role of Tet1 during gonadal

DNA demethylation in mouse primordial germ cells

3.1 Introduction

As discussed in detail in Chapter 1, gametes are highly specialised cells capable of fertilisation and supporting embryonic development of the next generation. In mouse, they are first specified in the developing embryo as primordial germ cells (PGCs) starting around embryonic day (E) 6.25 (Figure 1A; (Hill et al., 2014)). Following specification, the nascent PGCs undergo pronounced global epigenetic changes (de Napoles et al., 2007; Guibert et al., 2012; Hajkova et al., 2008; Kurimoto et al., 2015; Seisenberger et al., 2012; Seki et al., 2005; Seki et al., 2007; Sugimoto and Abe, 2007), including global reduction of genomic 5-methylcytosine (5mC; (Guibert et al., 2012; Seisenberger et al., 2012; Seki et al., 2005)). PGCs subsequently migrate from their location in the hindgut into the genital ridge where they undergo another wave of extensive epigenetic reprogramming starting at E10.5/E11.5 (Figure 1A; (Ancelin et al., 2006; Guibert et al., 2012; Hackett et al., 2013; Hajkova et al., 2008; Hajkova et al., 2002; Hajkova et al., 2010; Hill et al., 2014; Kagiwada et al., 2013; Lee et al., 2002; Popp et al., 2010; Seisenberger et al., 2012; Yamaguchi et al., 2012; Yamaguchi et al., 2013b)). Central to this wave of epigenetic reprogramming is the genome-wide loss of 5mC, including at imprint control regions (ICRs), germline gene promoters and most of the repetitive elements (Guibert et al., 2012; Hackett et al., 2013; Hajkova et al., 2008; Hajkova et al., 2002; Kagiwada et al., 2013; Lee et al., 2002; Seisenberger et al., 2012; Yamaguchi et al., 2012; Yamaguchi et al., 2013b).

The molecular mechanism(s) implicated in DNA demethylation of gonadal PGCs have been the focus of intense studies, with both active and passive mechanisms being proposed, as discussed in detail in Chapter 1 (Guibert et al., 2012; Hackett et al., 2013; Hajkova et al., 2008; Hajkova et al., 2002; Hajkova et al., 2010; Kagiwada et al., 2013; Lee et al., 2002; Popp et al., 2010; Seisenberger et al., 2012; Yamaguchi et al., 2013a; Yamaguchi et al., 2012; Yamaguchi et al., 2013b). Recently published observations have suggested that the 5mC oxygenase Tet1 constitutes a critical factor involved in the

correct progression of DNA demethylation in gonadal PGCs (Hackett et al., 2013; Hajkova et al., 2010; Yamaguchi et al., 2012; Yamaguchi et al., 2013b), with clearly detectable levels of 5-hydroxymethylcytosine (5hmC) observed at ICRs and germline gene promoters during epigenetic reprogramming (Hackett et al., 2013; Yamaguchi et al., 2013a). Additionally, hyper-methylation has been observed at these regions in *Tet1*-KO E13.5 PGCs (Yamaguchi et al., 2012; Yamaguchi et al., 2013b), *Tet1*-KO sperm (Dawlaty et al., 2013; Yamaguchi et al., 2013b), and the embryonic offspring of *Tet1*-KO males and females crossed with wild type animals (Yamaguchi et al., 2013b). However, the observed perturbations of DNA methylation appear to show a high degree of stochasticity, as the abnormalities affect only a subset of the analysed samples and genomic regions (Dawlaty et al., 2013; Yamaguchi et al., 2013b). These observations show a clear and surprising discrepancy considering the importance of germline reprogramming for resetting of genomic imprints in the gametes (Ferguson-Smith, 2011). Thus, to further clarify these observations, I set off to investigate in detail the functional relationship between DNA demethylation and Tet1-driven 5hmC formation in mouse gonadal PGCs.

3.2 Results

3.2.1 Re-evaluating 5hmC dynamics during gonadal DNA demethylation using integrative single nucleotide resolution DNA modification mapping and quantitative mass spectrometry

To first address the role of Tet1 during gonadal DNA demethylation in mouse primordial germ cells, I investigated the relationship between 5hmC and 5mC in wild type PGCs. Timing of gonadal DNA demethylation appears to be strain specific, varying up to two days depending on the genetic background of the mouse (Hackett et al., 2013; Hajkova et al., 2008; Hajkova et al., 2002; Seisenberger et al., 2012). In view of this, I first determined the dynamics of DNA demethylation in

PGCs isolated from embryos with an MF1 x GOF18ΔPE-EGFP (i.e. Oct4-GFP) genetic background. This background was chosen as it has previously been used to study 5hmC dynamics during gonadal epigenetic reprogramming in PGCs (by an antibody based hMeDIP approach) (Hackett et al., 2013), and thus offered the potential for direct comparison. To determine the dynamics of DNA demethylation in PGCs isolated from embryos with an MF1 x GOF18ΔPE-EGFP genetic background, I used two complementary approaches: 1) I used liquid-chromatography mass spectrometry (LC/MS, done in collaboration with Dr Rachel Amouroux) to determine global 5mC levels (relative to dG) for PGCs isolated from embryos between E9.5 and E13.5; and 2) I used whole genome bisulphite sequencing (WGBS) to study locus specific levels of total DNA modification (5mC + 5hmC) at cytosines in a CpG context in PGCs isolated from E10.5 to E12.5 embryos . Of note, between six and eight biological replicates per stage were used for LC/MS analysis (with DNA from 200-1,000 PGCs per sample used), and two biological replicates per stage were used for WGBS analysis (with DNA from 10,000 PGCs per sample used, which resulted in high (i.e.12-16x) genomic coverage per stage; see Table 3.1 for details). Additionally, for all analyses, male and female PGCs were treated separately from E12.5 onwards to account for any potential sex-specific differences.

Sample	Total aligned reads	Total aligned reads (de-duplicated)	Bisulfite conversion	CpGs covered (8x)	Mean genomic coverage	Mean methylation % (CpG context)	Mean methylation % (non-CpG context)
E10.5 PGCs (replicate 1)	213,778,056	106,299,921	NA	10,266,450	7.7x	19.6%	3.6%
E10.5 PGCs (replicate 2)	211,653,494	152,468,727	98%	11,076,503	8.2x	71.7%	3.0%
E11.5 PGCs (replicate 1)	217,727,331	156,802,841	NA	15,490,601	11.8x	9.3%	4.3%
E11.5 PGCs (replicate 2)	147,300,128	71,346,356	97%	1,243,506	3.6x	10.2%	4.0%
E12.5 female PGCs (replicate 1)	196,176,978	125,788,016	NA	12,984,269	9.7x	6.7%	5.3%
E12.5 female PGCs (replicate 2)	135,960,095	44,772,812	96%	141,201	2.2x	9.0%	5.5%
E12.5 male PGCs (replicate 1)	204,602,838	139,845,196	NA	14,013,444	10.7x	7.1%	4.7%
E12.5 male PGCs (replicate 2)	174,447,372	90,591,471	97%	3,201,912	4.6x	9.0%	3.5%

Table 3.1 – Summary of whole genome bisulphite sequencing (WGBS) statistics. *Where stated, bisulphite conversion represents the percentage of unconverted cytosines detected on spiked-in unmethylated λ phage DNA.*

As has previously been reported, DNA demethylation in PGCs of this genetic background is initiated starting at E10.5 (Figure 3.1A,C-D). Specifically, WGBS analysis revealed that near complete loss of total DNA modification occurs between E10.5 and E11.5 at promoters (TSS -1kB/+500bp), gene bodies, intergenic regions, putative active enhancers (as determined by the presence of the specific $H3K27Ac^{+}/H3K4me3^{-}$ chromatin signature in day 6 PGC-like cells (Kurimoto et al., 2015)) and

imprint control regions (Figures 3.1C, 3.5D). Further careful analysis of WGBS biological replicates for PGCs isolated from E10.5 to E12.5 embryos suggested that DNA demethylation at regions within the uniquely mappable part of the genome (i.e. promoters, gene bodies, etc) may already be underway at E10.5 (Figure 3.1C). Specifically, total DNA modification levels for E10.5 biological replicate 1 (E10R1) at these particular loci are clearly intermediate between E10.5 biological replicate 2 (E10R2) and both E11.5 samples (Figure 3.1B-C); in this context, it should be noted that the large difference in total DNA modification between E10.5 biological replicates cannot be explained by incomplete bisulphite conversion of the E10R2 DNA during WGBS library preparation (98% conversion; Table 3.1).

While WGBS analysis suggested that DNA demethylation at regions within the uniquely mappable part of the genome is initiated on E10.5, LC/MS analysis surprisingly suggested that global DNA demethylation is initiated between E10.5 and E11.5, rather than on E10.5 itself (Figure 3.1A). Furthermore, while LC/MS analysis confirmed that the major drop in global DNA methylation levels is observed between E10.5 and E11.5, which would be consistent with WGBS analysis (Figure 3.1C), LC/MS analysis also suggested that DNA demethylation is clearly still ongoing between E11.5 and E13.5 (Figure 3.1A). In contrast, WGBS analysis suggested that DNA demethylation at features within the uniquely mappable part of the genome appears to have largely completed by E11.5, with only limited DNA demethylation observed between E11.5 and E12.5 (Figure 3.1C). These discrepancies may be explained in part by the fact that LC/MS predominantly reflects methylation at repetitive elements (which, despite making up approximately 40% of the genome, contain the bulk of 5mC), while WGBS reflects predominantly methylation levels at loci contained within the uniquely mappable part of the genome (thus under-representing repeats). Indeed, upon aligning WGBS reads to consensus sequences for various repetitive elements, it is clear that the bulk of DNA demethylation occurs between E10.5 and E11.5, rather than on E10.5 itself, at highly-methylated high copy number repeats such as LINE1, IAP and satellite elements (Figure 3.1D). Moreover, DNA demethylation is clearly still ongoing between E11.5 and E12.5 at these regions (Figure 3.1D). In contrast, other

repetitive elements, such as MuLV and Etn1 ERVs have already completed their demethylation by E11.5, similar to the uniquely mappable parts of the genome (Figure 3.1C-D). Thus, the timing of DNA demethylation across the genome does not appear uniform in mouse PGCs, with the uniquely mappable part of the genome demethylating earlier and faster than certain high copy-number repeats.

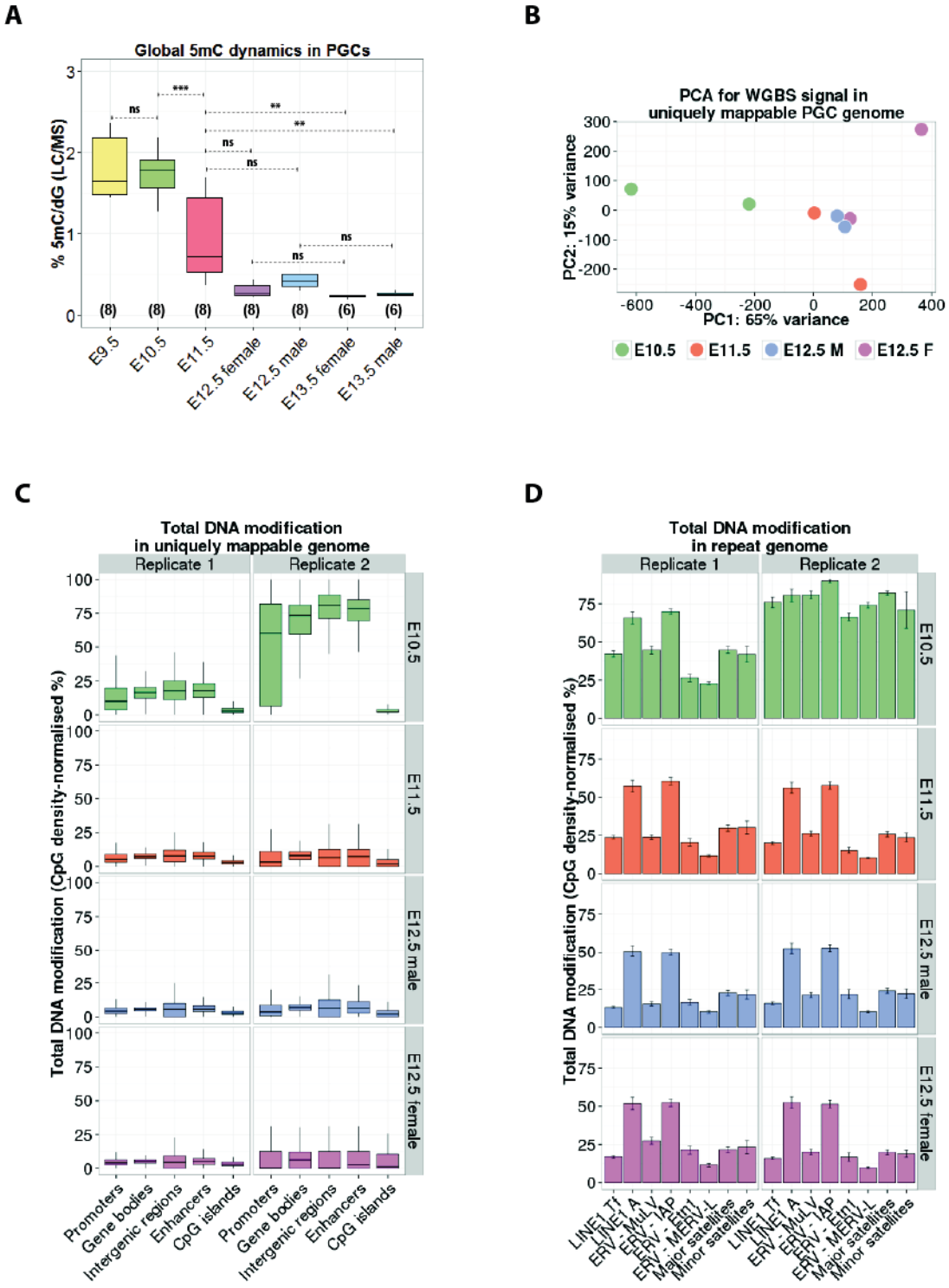


Figure 3.1 – Characterisation of gonadal DNA demethylation in PGCs derived from MF1 x GOF18ΔPE-EGFP embryos. A) *Total 5mC levels between E9.5 and E13.5 in wild type PGCs as measured by LC/MS. p-values based on combined ANOVA and Tukey post hoc test; ***: p-value<0.001, **: p-value<0.01, *: p-value<0.05. The number of biological replicates measured for each stage is indicated below the box.* ***B)*** *Principle component analysis showing clustering of WGBS biological replicates between E10.5 and E12.5. PC1 and PC2 refer to principle components 1 and 2, respectively.* ***C)*** *DNA methylation levels in biological replicates between E10.5 and E12.5 for various features within the uniquely mappable genome.* ***D)*** *DNA methylation levels in biological replicates between E10.5 and E12.5 for various repetitive elements. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge. Note that for all bar charts, bars represent mean value and error bars represent standard error of the mean.*

I next sought to directly address the relationship between 5hmC dynamics and DNA demethylation in PGCs. To first address this, I carried out quantitative LC/MS analysis on PGCs between E9.5 and E13.5. This revealed relatively constant 5hmC levels across these stages (Figure 3.5A), with a very slight increase (13%) over base line levels observed at E11.5 and a slight decrease in 5hmC levels observed in females starting at E12.5. Importantly, 5hmC levels are consistently an order of magnitude lower (0.09% 5hmC/dG at E11.5) (Figure 3.5A) than either the total 5mC levels observed at E10.5 (1.73% 5mC/dG) (Figure 3.1A) or the amount of 5mC lost between E10.5 and E11.5 (0.82% 5mC/dG) (Figure 3.1A). Thus, DNA demethylation is not accompanied by a reciprocal increase in 5hmC levels, as would be predicted for a mechanism whereby oxidation of 5mC to 5hmC followed by passive dilution of the oxidised base is primarily responsible for driving global DNA demethylation (Hackett et al., 2013). Considering these observations, quantification of global DNA modification levels in wild type PGCs hinted at a more complex relationship between 5hmC and DNA demethylation than has been previously proposed (Hackett et al., 2013).

To expand upon these initial findings, I carried out a detailed analysis of genome-wide locus-specific DNA modification dynamics during gonadal DNA demethylation by complementing the

aforementioned WGBS analysis in PGCs between E10.5 and E12.5 with single-nucleotide resolution 5hmC mapping. Traditionally, genome wide mapping of 5hmC is carried out either by antibody based immuno-precipitation enrichment methods (e.g. hMeDIP, using a 5hmC-specific antibody) (Hackett et al., 2013; Pastor et al., 2011; Robertson et al., 2011; Song et al., 2011) or by single nucleotide resolution mapping methods (Booth et al., 2012; Yu et al., 2012). The most commonly used single nucleotide resolution 5hmC mapping approach (Tet-assisted bisulphite sequencing or TAB-Seq) is a three-step procedure: 1) 5hmC is enzymatically modified to prevent further oxidation; 2) 5mC is oxidised to 5caC by Tet catalytic activity; and finally 3) traditional whole genome bisulphite sequencing is carried out on the end product (Yu et al., 2012). As 5caC and unmodified cytosine both read as thymine following bisulphite treatment and PCR amplification while (modified) 5hmC reads as cytosine, 5hmC-specific sites can be distinguished in this way at base-pair resolution (Yu et al., 2012).

The current antibody-based and single-nucleotide resolution methods, however, have non trivial disadvantages, especially when working with precious samples where only a limited amount of DNA can be isolated (such as PGCs). Antibody-based detection has been shown to exhibit bias toward epitope-dense regions (Yu et al., 2012); specifically, CpG rich regions (i.e. CpG islands) with relatively low 5hmC levels can easily appear highly hydroxy-methylated compared to CpG-poor regions with relatively high 5hmC levels simply due to a high local concentration of epitope in the CpG-rich regions. Thus, the interpretation of antibody based approaches is non-trivial. On the other hand, single-nucleotide resolution methods do not suffer from these biases. However, the most commonly used single nucleotide resolution method (TAB-Seq) requires complete oxidation of 5mC to 5caC enzymatically using Tet proteins, and thus TAB-Seq requires highly active Tet enzymes to obtain a high conversion rate of 5mC to 5caC (Yu et al., 2012).

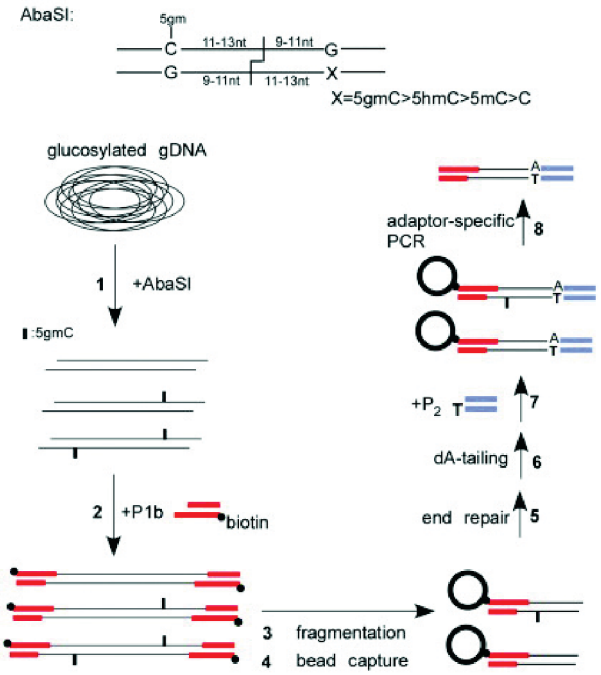


Figure 3.2 – Overview of AbaSeq method - adapted from (Sun et al., 2013).

Considering the above, we collaborated with Dr Shriharsa Pradhan and colleagues at New England Biolabs (Ipswich, MA, USA), who had previously developed and validated a quantitative single-nucleotide resolution 5hmC mapping method using a new enzymatic approach optimised for low DNA input (Sun et al., 2013). In this method, called AbaSeq, 5hmC on genomic DNA is first enzymatically marked by addition of a glucosyl group by the T4-β-glucosyltransferase to generate glucosylated-5hmC (5gmC) (Figure 3.2). Glucosylated DNA is then digested by the AbaSI restriction endonuclease, which recognizes 5gmC with high specificity and generates double-stranded breaks at the 3′ side of the binding sites 11–13 nucleotides away in the top strand and 9–10 nucleotides away in the bottom strand (Figure 3.2) (Sun et al., 2013). The AbaSI-digested DNA is subsequently ligated to a biotinylated P1b DNA adaptor with a 2-base randomized 3′-overhang and further fragmented by random shearing (Figure 3.2) (Sun et al., 2013). Finally, next generation sequencing libraries are prepared from the smaller purified biotinylated fragments (Figure 3.2). Following sequencing from the P1b adapter end and alignment to a reference genome, the neighbouring sequences of the aligned fragments are interrogated in search of putative 5hmC sites at the expected distance (11–13 nucleotides away in the top strand and 9–10 nucleotides away in the bottom strand) (Sun et al., 2013).

In this way, the relative 5hmC levels at each cytosine can be determined based on the total number of reads mapped back to the site. Upon normalisation to the total number of reads in a library and the global 5hmC levels, as determined by quantitative mass spectrometry, AbaSeq can also be used to directly compare site-specific 5hmC levels in different samples. Of note, it should be mentioned that this method has been extensively validated in mouse embryonic stem cells (Sun et al., 2013).

Sample	Mappable reads	Reads with multiple hits	Uniquely mapped reads
E10.5 PGCs (replicate 1)	183,711,239	28,715,048	154,996,191
E10.5 PGCs (replicate 2)	178,068,245	36,937,858	141,130,387
E11.5 PGCs (replicate 1)	98,902,531	37,981,359	60,921,172
E11.5 PGCs (replicate 2)	167,669,918	64,751,520	102,918,398
E12.5 female PGCs (replicate 1)	111,003,663	62,975,795	48,027,868
E12.5 female PGCs (replicate 2)	178,561,029	91,371,336	87,189,693
E12.5 male PGCs (replicate 1)	43,926,295	29,574,693	14,351,602
E12.5 male PGCs (replicate 2)	164,194,628	94,643,413	69,551,215

Table 3.2 – Summary of AbaSeq sequencing statistics

Importantly, initial AbaSeq data analysis on samples derived from E10.5 to E12.5 PGCs (10,000 PGCs per sample; see Table 3.2 for details) confirmed reproducibility of the method, with significant correlation observed between biological replicates (Figure 3.3A). Additionally, principal component analysis revealed clustering of samples by stage along principal component 1, which contributes 76% of the total variance observed between samples (Figure 3.3B). Moreover, I observed a high correlation between total 5hmC levels at promoter regions (-1kB/+500bp) in E10.5 PGCs as computed by AbaSeq and as computed by a previous antibody-based hMeDIP approach (Figure 3.1C) (Hackett et al., 2013). It should be noted that, upon normalisation of AbaSeq signal to CpG density (which is only possible using a single nucleotide resolution method), the correlation between AbaSeq signal and hMeDIP signal decreases considerably, with a small subpopulation of promoters showing low mean 5hmC signal in AbaSeq but high total 5hmC signal in hMeDIP (Figure 3.3D); a similar small subpopulation of promoters is also observed when I compared CpG-normalised and total 5hmC signal at E10.5 PGC promoters as computed by AbaSeq (Figure 3.3E). This discrepancy between total and CpG density-normalised 5hmC levels appears to be due to the over-estimation of 5hmC signal at high-CpG content promoters and the under-estimation of 5hmC signal at low-CpG content promoters

(Figure 3.3F), highlighting the previously discussed bias associated with antibody-based 5hmC mapping methods. Of note, a similar bias is observed when comparing antibody based MeDIP and single-nucleotide resolution WGBS approaches to compute 5mC (MeDIP) (Hackett et al., 2013) and total DNA modification (5mC+ 5hmC; WGBS) levels in E10.5 PGCs (Figures 3.3G-H). Specifically, promoter methylation as determined by MeDIP and WGBS only have a Spearman correlation of 0.42 and a Pearson correlation of 0.02 (Figures 3.3G-H); by comparison, the two WGBS biological replicates, despite one already showing considerable demethylation at promoter regions, have a Spearman correlation of 0.79 and a Pearson correlation of 0.74. Although WGBS signal represents the combined 5mC and 5hmC levels, as bisulphite sequencing cannot distinguish between these two modifications (Huang et al., 2010), it remains appropriate to use for direct comparison with MeDIP datasets considering that total 5mC levels are approximately 20-fold higher than total 5hmC levels in PGCs at E10.5 (Figures 3.1A, 3.5A), and thus the vast majority of WGBS signal will be 5mC. Together, these observations draw attention to the power of single-nucleotide resolution methods for mapping DNA modifications over antibody based approaches.

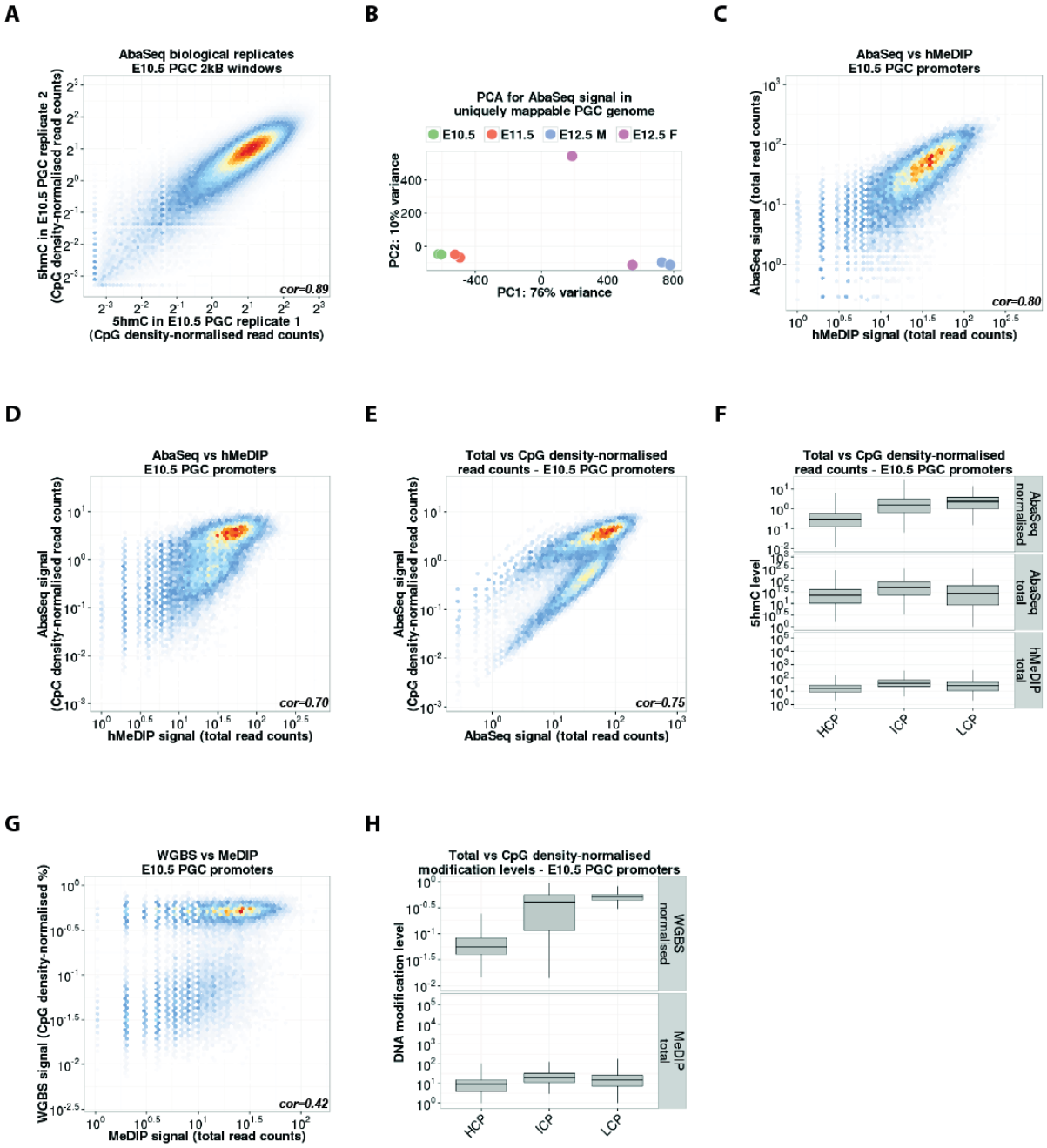


Figure 3.3 – Validation of AbaSeq method in PGCs. A) *Density heatmap showing mean 5hmC levels per 2kB window (with minimum 4 CpGs) for E10.5 PGC replicate 1 (x-axis) and replicate 2 (y-axis). Spearman correlation coefficient is shown.* ***B)*** *Principle component analysis showing clustering of AbaSeq biological replicates between E10.5 and E12.5.* ***C)*** *Density heatmap showing total 5hmC levels at promoter regions (-1kB/+500bp) in E10.5 PGCs as computed by AbaSeq (y-axis) and hMeDIP (x-axis,(Hackett et al., 2013)). Spearman correlation coefficient is shown.* ***D)*** *Density heatmap showing AbaSeq-computed CpG density-normalised 5hmC levels (y-axis) and hMeDIP-computed total 5hmC levels (x-axis, (Hackett et al., 2013)) at promoter regions (-1kB/+500bp) in E10.5 PGCs. Spearman correlation coefficient is shown.* ***E)*** *Density heatmap showing AbaSeq-computed CpG density-normalised 5hmC levels (y-axis) and AbaSeq-computed total 5hmC levels (x-axis,(Hackett et al., 2013)) at promoter regions (-1kB/+500bp) in E10.5 PGCs. Spearman correlation coefficient is shown.* ***F)*** *AbaSeq-computed CpG density-normalised 5hmC levels (top panel), AbaSeq-*

computed total 5hmC levels (middle panel), and hMeDIP-computed total 5hmC levels (lower panel) for high-CpG (HCP), intermediate CpG (ICP) and low-CpG (LCP) promoters (as defined in (Borgel et al., 2010)) in E10.5 PGCs. ***G)*** *Density heatmap showing WGBS-computed CpG density-normalised total DNA modification levels (y-axis) and MeDIP-computed total 5mC levels (x-axis, (Hackett et al., 2013)) at promoter regions (-1kB/+500bp) in E10.5 PGCs. Spearman correlation coefficient is shown.* ***H)*** *WGBS-computed CpG density-normalised total DNA modification levels (top panel) and MeDIP-computed total 5mC levels (bottom panel, (Hackett et al., 2013)) for high-CpG (HCP), intermediate CpG (ICP) and low-CpG (LCP) promoters (as defined in (Borgel et al., 2010)) in E10.5 PGCs. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.*

With respect to the relationship between 5hmC dynamics and DNA demethylation in PGCs, I first carried out an in detail analysis of 5hmC localisation in E10.5 PGCs (Figure 3.4), at which point gonadal DNA demethylation is initiated (Figure 3.1). Surprisingly, 5hmC localisation at this stage is remarkably similar to the 5hmC deposition patterns observed in mouse embryonic stem cells (mESCs) (Figures 3.4A,B) (Sun et al., 2013; Yu et al., 2012), a cell system which notably does not undergo genome-wide DNA demethylation when cultured in serum-containing media. This similarity is most strikingly highlighted by the strong (Spearman correlation = 0.73) positive correlation observed between 5hmC levels at all 2kB windows genome wide in E10.5 PGCs and in mESCs (Figure 3.4B). More generally, in E10.5 PGCs, 5hmC levels are slightly lower than the genomic average at promoter regions (Figures 3.4A,D); in line with genomic average at gene bodies and intergenic loci (Figures 3.4A,D); depleted within CpG islands (Figures 3.4A); and highly enriched at putative active enhancers (Figures 3.4A,C). Moreover, 5hmC deposition at promoter regions shows a clear inverse relationship with gene expression levels (as determined by RNA-Seq; see Chapter 5 for details), similar to the patterns observed for total DNA modification (Figure 3.4D). Within gene bodies, on the other hand, 5hmC is clearly enriched at expressed genes when compared with the levels observed for genes without detectable expression (Figure 3.4D). Intriguingly, however, a non-linear relationship with gene expression is observed (Figure 3.4D), with the highest 5hmC levels detected at median and low expressed genes; by comparison, WGBS signal across gene bodies shows a clear positive relationship with gene expression levels (Figure 3.4D). Of note, the relationship between gene

expression levels and DNA modification levels hold true for both AbaSeq and WGBS biological replicates.

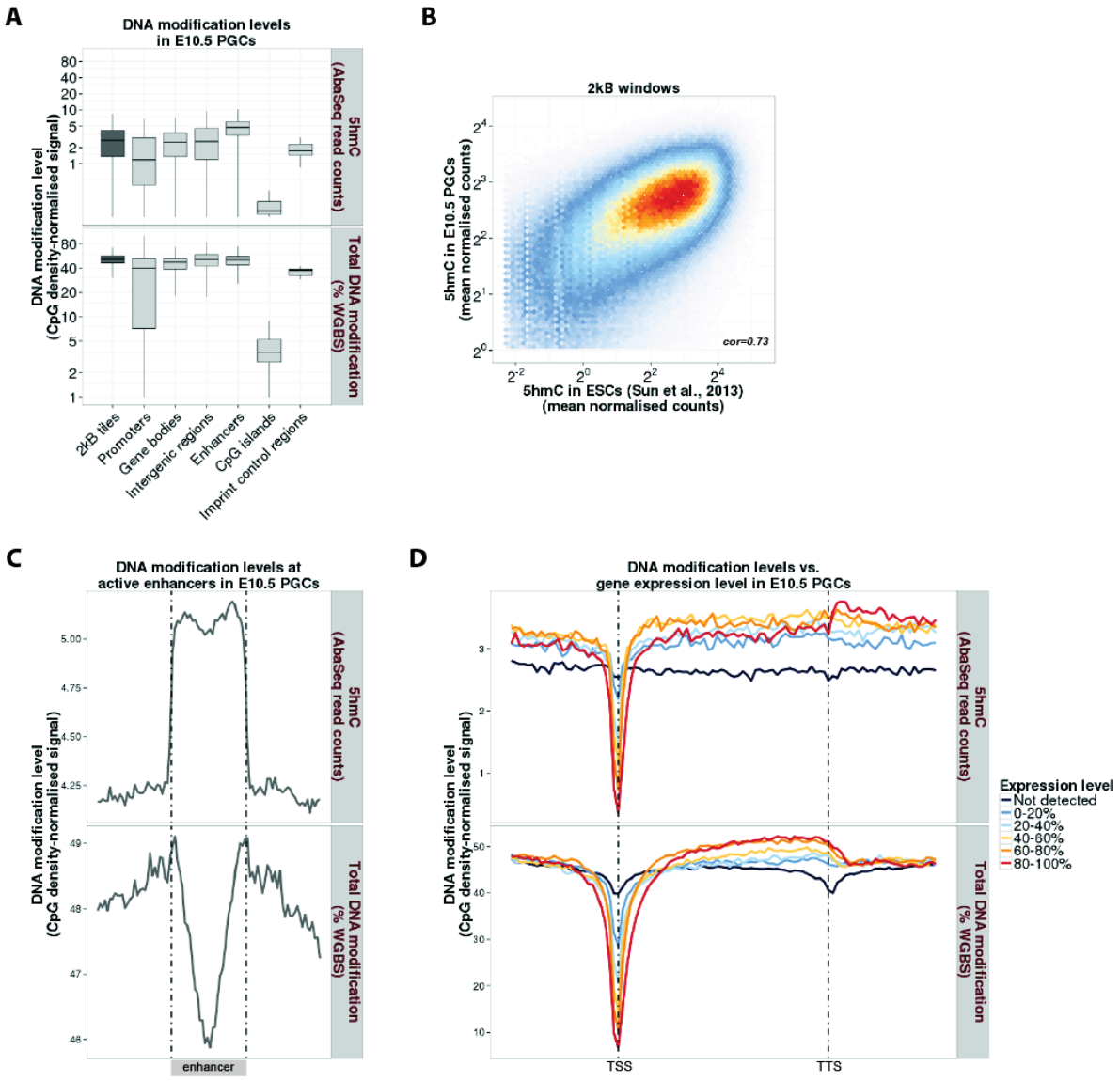


Figure 3.4 – 5hmC localisation in E10.5 PGCs. A) *5hmC levels (top panel) and total DNA modification levels (bottom panel) at various regulatory elements in E10.5 PGCs.* ***B)*** *Density heatmap showing mean 5hmC levels per 2kB window (with minimum 4 CpGs) for mouse embryonic stem cells (x-axis, (Sun et al., 2013)) and E10.5 PGCs (y-axis). Spearman correlation coefficient is shown.* ***C)*** *Metagene plot showing 5hmC levels (top panel) and total DNA modification level (bottom panel) across putative active enhancers.* ***D)*** *Metagene plot showing 5hmC levels (top panel) and total DNA modification level (bottom panel) across genes expressed at different levels in E10.5 PGCs. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.*

To expand upon these initial findings, I carried out a detailed analysis of genome-wide locus-specific DNA modification dynamics during gonadal DNA demethylation. Intriguingly, this revealed that 5hmC dynamics are not as stable as suggested from LC/MS analysis; rather, while global levels of 5hmC remain constant between E10.5 and E12.5, 5hmC is re-localised during this period from the uniquely mappable part of the genome to repetitive elements in both males and females (Figures 3.5B-C). As such, 5hmC levels decrease at regulatory elements, gene bodies, and intergenic regions between E10.5 and E12.5 (Figure 3.5D), concurrent with a general increase in levels at satellite repeats and retrotransposons (Figure 3.5E). This 5hmC re-localisation is also clearly observable by immunofluorescence staining of 5hmC in PGCs (Figure 3.5C). While 5hmC loss throughout the uniquely mappable part of the genome matches the loss of total DNA modification, it should be noted that this does not occur with the same kinetics (Figure 3.5D). Specifically, the vast majority of WGBS signal is lost between E10.5 and E11.5 in the uniquely mappable part of the genome (Figure 3.5D); in contrast, 5hmC signal generally persists at these regions until E11.5, only to be subsequently lost between E11.5 and E12.5 (Figure 3.5D). Thus, although the absolute amount of 5hmC in the genome as determined by LC/MS remains largely unchanged (Figure 3.5A), 5hmC represents a considerably higher proportion of total DNA modification in the uniquely mappable part of the genome in PGCs at E11.5, immediately following DNA demethylation, compared with the preceding (E10.5) or following (E12.5) stages (Figure 3.5F).

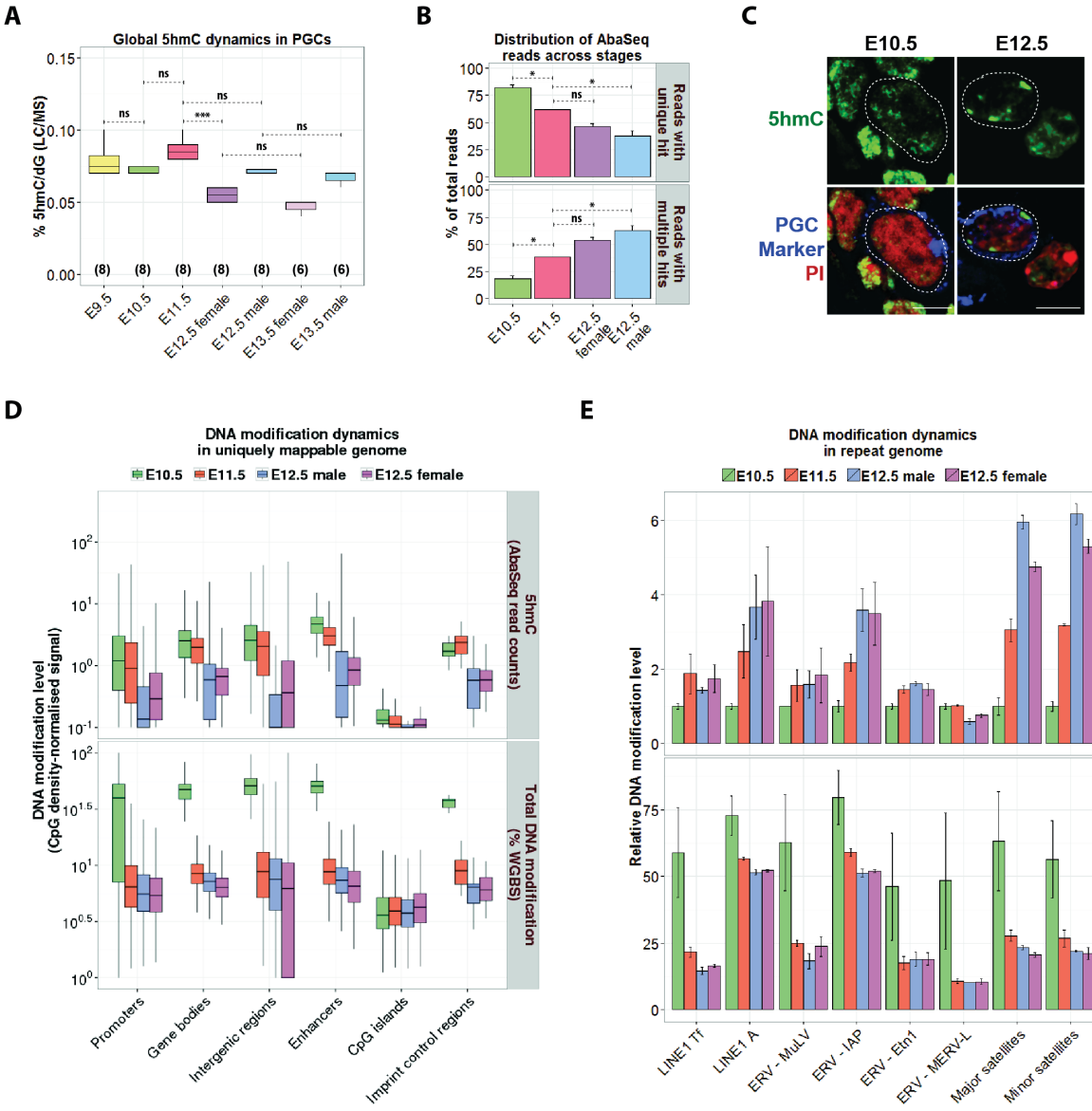
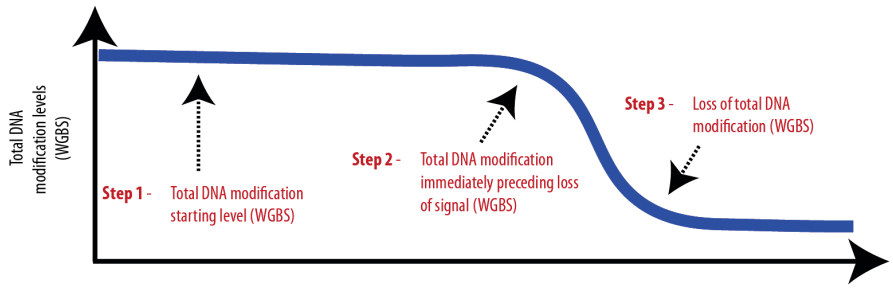

A
Global 5hmC dynamics in PGCs
% 5hmC/dG (LC/MS)
E9.5
E10.5
E11.5
E12.5 female
E12.5 male
E13.5 female
E13.5 male
B
Distribution of AbaSeq reads across stages
% of total reads
Reads with unique hit
Reads with multiple hits
C
E10.5
E12.5
5hmC
PGC Marker
PI
D
DNA modification dynamics in uniquely mappable genome
E10.5 E11.5 E12.5 male E12.5 female
DNA modification level (CpG density-normalised signal)
5hmC (AbaSeq read counts)
Total DNA modification (% WGBS)
Promoters
Gene bodies
Intergenic regions
Enhancers
CpG islands
Imprint control regions
E
DNA modification dynamics in repeat genome
E10.5 E11.5 E12.5 male E12.5 female
Relative DNA modification level
5hmC (AbaSeq reads) (normalised to E10.5)
Total DNA modification (% WGBS)
LINE1 Tf
LINE1 A
ERV - MuLV
ERV - IAP
ERV - Etn1
ERV - MERV-L
Major satellites
Minor satellites
F
Uniquely mapped 2kB windows
DNA modification levels (CpG density-normalised signal)
Total DNA modification (% WGBS)
5hmC enrichment (AbaSeq/WGBS)
E10.5
E11.5
E12.5 male
E12.5 female

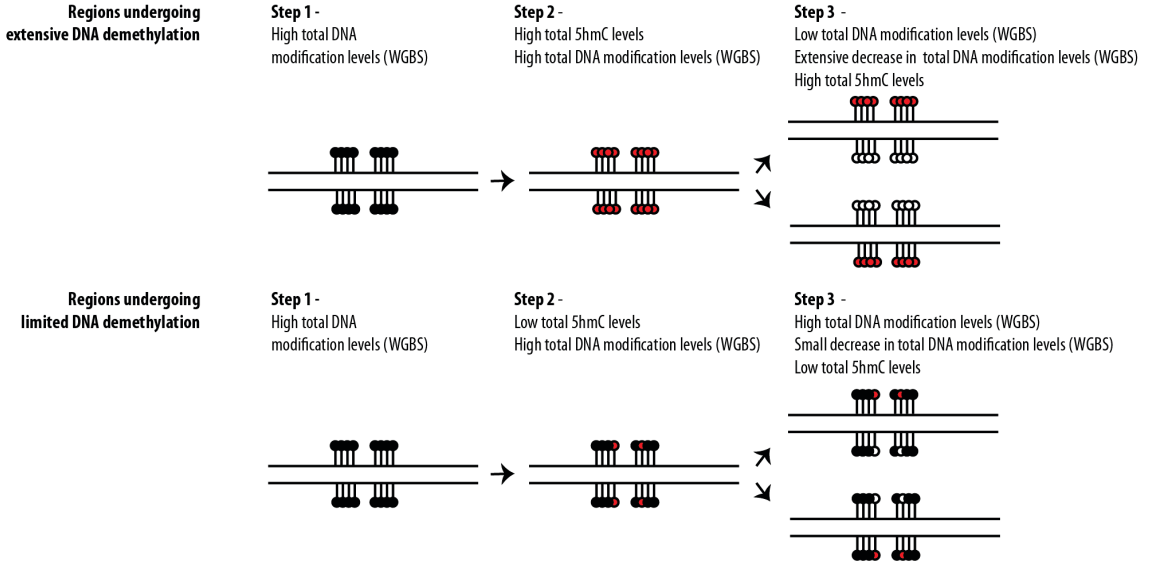
Figure 3.5 – Global and genome-wide locus-specific 5hmC dynamics during gonadal DNA demethylation in PGCs. A) *Global 5hmC levels between E9.5 and E13.5 in wild type PGCs as determined by LC/MS. p-values based on combined ANOVA and Tukey post hoc test; ***: p-value<0.001, **: p-value<0.01, *: p-value<0.05. The number of biological replicates measured for each stage is indicated below the box.* ***B)*** *Re-distribution of 5hmC from the uniquely mappable part of the genome to repetitive elements between E10.5 and E12.5. p-values based on combined ANOVA and Tukey post hoc test; ***: p-value<0.001, **: p-value<0.01, *: p-value<0.05.* ***C)*** *Immunostaining for 5hmC in E10.5 and E12.5 PGCs. PGCs are marked in blue by staining for either SSEA-1 (E10.5) or Ddx4 (E12.5). Note the diffuse staining pattern for 5hmC in E10.5 PGCs compared with the focal staining in E12.5 PGCs. Scale bar represents 10 µm.* ***D)*** *Relative 5hmC (top panel) and total DNA modification (bottom panel) levels at various features within the uniquely mappable part of the genome between E10.5 and E12.5.* ***E)*** *Relative 5hmC (top panel) and total DNA modification (bottom panel) levels at various classes of repetitive elements between E10.5 and E12.5. Note that 5hmC level for each repeat at each stage is normalised to mean 5hmC level at E10.5.* ***F)*** *Relative enrichment for 5hmC over total DNA modification in PGCs between E10.5 and E12.5 across all 2kB windows within the uniquely mappable part of the genome. p-values computed by Mann-Whitney U-test. p-values based on combined ANOVA and Tukey post hoc test; ***: p-value<0.001, **: p-value<0.01, *: p-value<0.05. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge. Note that for all bar charts, bars represent mean value and error bars represent standard error of the mean.*

The offset kinetics regarding the loss of total DNA modification and loss of 5hmC at regions within the uniquely mappable part of the genome suggested that a model of oxidation followed by passive dilution may be functionally implicated in DNA demethylation specifically at these regions (Figure 3.6A). Alternatively, the observations described above are equally consistent with a model whereby oxidation of 5mC at loci within the uniquely mappable part of the genome may be involved in initiating DNA demethylation at these regions via an active mechanism (Figure 3.6B). In addition to a potential role in initiating DNA demethylation, a recent study from our lab has implicated Tet3-catalysed 5mC oxidation in protecting against aberrant *de novo* methylation in the zygote *following* DNA demethylation of the paternal pronucleus (Amouroux et al., 2016). I thus additionally considered that oxidation of 5mC to 5hmC (and its higher oxidative derivatives) at de-methylated regions within the uniquely mappable part of the genome may similarly help to maintain a hypo-methylated state following the major loss of 5mC in PGCs (Figure 3.6C). To distinguish between these three hypothesized models I considered the various predictions made by each one. A model of

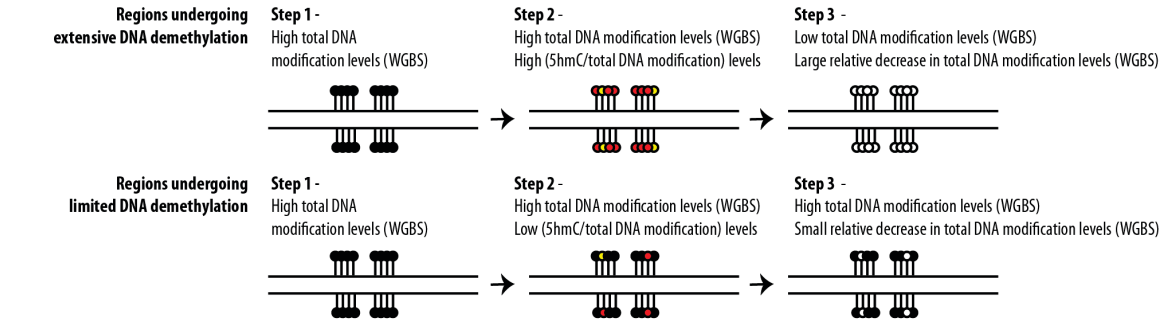
oxidation followed by passive dilution predominantly driving DNA demethylation at regions within the uniquely mappable part of the genome (Figure 3.6A) predicts an obvious positive correlation between the total amount of total DNA modification lost (i.e. 5mC + 5hmC, as determined by WGBS) between two stages and the total 5hmC levels at both the stage immediately preceding and following the decrease. Alternatively, a model of 5mC oxidation triggering active DNA demethylation (Figure 3.6B) predicts an obvious positive correlation between the relative extent to which total DNA modification levels decrease (i.e. based on the fold-reduction coefficient) between two stages at a given region and the relative 5hmC levels (i.e. 5hmC/(5mC + 5hmC)) at that region in the immediately preceding stage, as further oxidation of 5hmC to 5fC is the rate limiting step in the full oxidation of 5mC to 5caC (Hashimoto et al., 2014). Finally, a model implicating oxidation of 5mC in maintaining the hypo-methylated state following DNA demethylation (Figure 3.6C) predicts an obvious positive correlation between the relative extent to which total DNA modification levels decrease between two stages at a genomic interval and the relative 5hmC levels at that interval in the later of the two stages.



A Predictions for a model implicating 5hmC in passive DNA demethylation



B Predictions for a model implicating 5mC oxidation in initiation of active DNA demethylation



C Predictions for a model implicating 5mC oxidation in retaining the hypomethylated DNA state

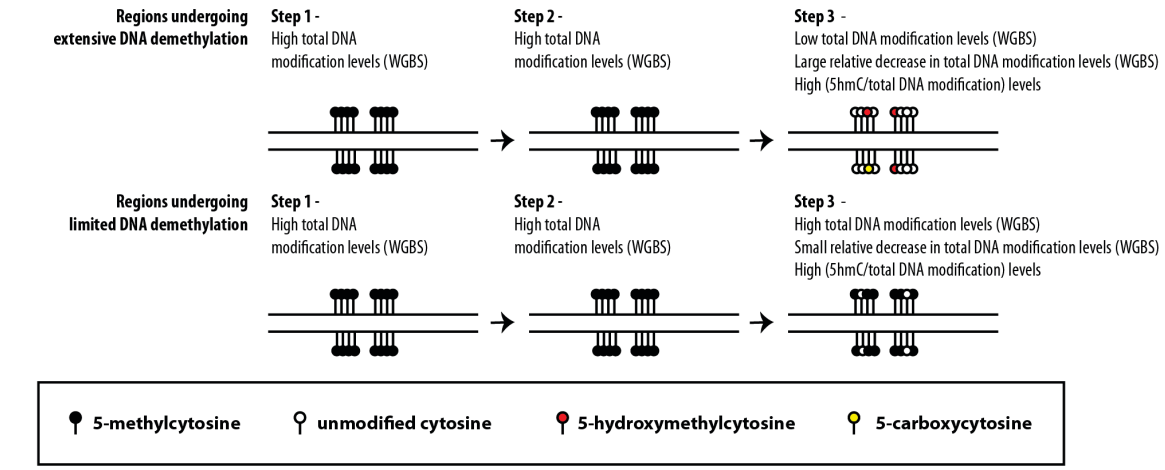


Figure 3.6 – Predictions made by each potential model implicating 5mC oxidation in DNA demethylation of gonadal PGCs. A) *A model of oxidation followed by passive dilution predominantly driving DNA demethylation predicts a positive correlation between the extent to which the combined 5mC/5hmC levels decrease between two stages (i.e. %, WGBS) and the total level of 5hmC at both the stage immediately preceding and following this decrease.* ***B)*** *A model implicating 5mC oxidation in initiating DNA demethylation via an active mechanism predicts a positive correlation between the extent to which the combined 5mC/5hmC levels decrease between two stages (i.e. %, WGBS) and the relative 5hmC levels in the stage immediately preceding this decrease, as further oxidation of 5hmC to 5fC is the rate limiting step in the full oxidation of 5mC to 5caC (Hashimoto et al., 2014).* ***C)*** *A model implicating oxidation of 5mC in maintaining a hypo-methylated DNA state following the major wave of DNA demethylation predicts that regions where the majority of DNA demethylation has been lost between two stages (i.e. are newly hypo-methylated) will have high relative levels of 5hmC in the stage immediately following the major wave of DNA demethylation to remove residual methylation and/or aberrant de novo methylation. Thus, a limited correlation between the extent to which the combined 5mC/5hmC levels decrease between two stages (i.e. %, WGBS) and the relative 5hmC levels in the stage immediately following this decrease may also be seen.*

To test each prediction, I carried out pairwise correlation analysis between either the total or relative amount of total DNA modification (i.e. 5mC + 5hmC, as determined by WGBS) lost between two given stages and either the total or relative 5hmC levels for all 2kB windows with a minimum 40% total DNA modification in E10.5 PGCs (Figure 3.7A). Additionally, all correlations were done using 5hmC levels from both the stage preceding and following loss of total DNA modification. Remarkably, by far the largest positive correlation computed (Spearman correlation=0.58) after testing all potential relationships was between the relative amount of total DNA modification lost during the major wave of DNA demethylation at uniquely mapped loci (E10.5 to E11.5 PGCs) and the relative enrichment of 5hmC in the stage *following* this decrease (E11.5) (Figures 3.7A,G). Consistently, with respect to each stage, the largest positive correlation computed (E10.5 to E11.5, Spearman correlation=0.58; E11.5 to E12.5 (female), Spearman correlation=0.42; E11.5 to E12.5 (male), Spearman correlation=0.15) was always between the relative amount of total DNA modification lost between two stages and the relative enrichment of 5hmC in the stage *following* the decrease (Figure 3.7A). It should be noted that this correlation is clearly weaker when studying the relationship between the relative amount of total DNA modification lost from E11.5 to E12.5 and the relative 5hmC levels in E12.5 PGCs, which likely reflects the fact that only limited DNA demethylation takes place at regions within the uniquely mappable part of the genome between these

stages (Figures 3.5D, 3.7A) and that the majority of 5hmC is localised to repetitive elements by E12.5 (Figures 3.5,B,C,E) in PGCs of this particular genetic background.

Considering the previously proposed role for 5hmC as an integral part of gonadal DNA demethylation, it was of particular note that no obvious correlation was observed between either the total or relative amount of total DNA modification (i.e. 5mC + 5hmC) lost between two stages (at all 2kB windows with a minimum of 40% total DNA modification at E10.5) and either the total or relative (i.e. 5hmC/(5mC + 5hmC)) 5hmC levels in the stage immediately preceding this decrease (Figures 3.7A-D). While this does not directly exclude a role for 5mC oxidation in DNA demethylation, in view of the much stronger correlation discussed above between the relative amount of total DNA modification lost between two stages and the relative enrichment of 5hmC in the stage *following* this decrease, these observations suggest that the role of 5mC oxidation during gonadal DNA demethylation in PGCs may be most functionally important following the major loss of 5mC either: 1) to protect newly hypo-methylated regions from sporadic DNA methylation; or 2) to clean-up residual DNA methylation that remains at these now largely de-methylated loci. Consistent with these hypotheses, regions depleted of 5hmC in both E10.5 and E11.5 PGCs are still able to undergo considerable DNA demethylation between these two stages (median 39% reduction in total DNA modification levels; median 4.5 fold reduction) (Figure 3.7H). However, regions enriched for 5hmC at either of these stages are slightly more hypo-methylated at E11.5 than regions depleted of 5hmC at both time points (Figure 3.7I). Again, this would be consistent with a model whereby DNA demethylation proceeds independent of 5hmC, but subsequent Tet catalytic activity helps to maintain a de-methylated state. Moreover, the relative 5hmC level (i.e. 5hmC/(5mC+5hmC)) at E11.5 for all 2kB windows (with a minimum 40% total DNA modification level at E10.5) clearly shows a negative correlation with the total amount of DNA modification remaining at this stage (Figure 3.7J), thus suggesting that 5hmC is targeted to newly hypo-methylated regions. Altogether, a model whereby oxidation of 5mC to 5hmC (and its higher oxidative derivatives) maintains a hypo-methylated state following the major loss of 5mC (Figure 3.6C) best fits the combined LC/MS and single-nucleotide

resolution mapping analyses I present above. I thus set out to validate such a model using gene knockout studies.

A

Samples used for determining extent of WGBS signal loss	Absolute reduction in WGBS signal (median value)	Fold reduction in WGBS signal (median value)	Samples used for determining 5hmC levels and enrichment	Correlation - absolute WGBS signal loss vs total 5hmC levels	Correlation - relative WGBS signal loss vs total 5hmC levels	Correlation - relative WGBS signal loss vs relative 5hmC enrichment
E10.5 PGCs/E11.5 PGCs	42.6%	5.9	E10.5 PGCs	0.14	0.22	0.21
E10.5 PGCs/E11.5 PGCs	42.6%	5.9	E11.5 PGCs	0.14	0.06	0.58
E11.5 PGCs/E12.5 female PGCs	2.6%	1.4	E11.5 PGCs	0.05	0.07	-0.23
E11.5 PGCs/E12.5 female PGCs	2.6%	1.4	E12.5 female PGCs	0.08	0.04	0.42
E11.5 PGCs/E12.5 male PGCs	1.6%	1.2	E11.5 PGCs	0.04	0.05	-0.31
E11.5 PGCs/E12.5 male PGCs	1.6%	1.2	E12.5 male PGCs	0.08	0.05	0.16

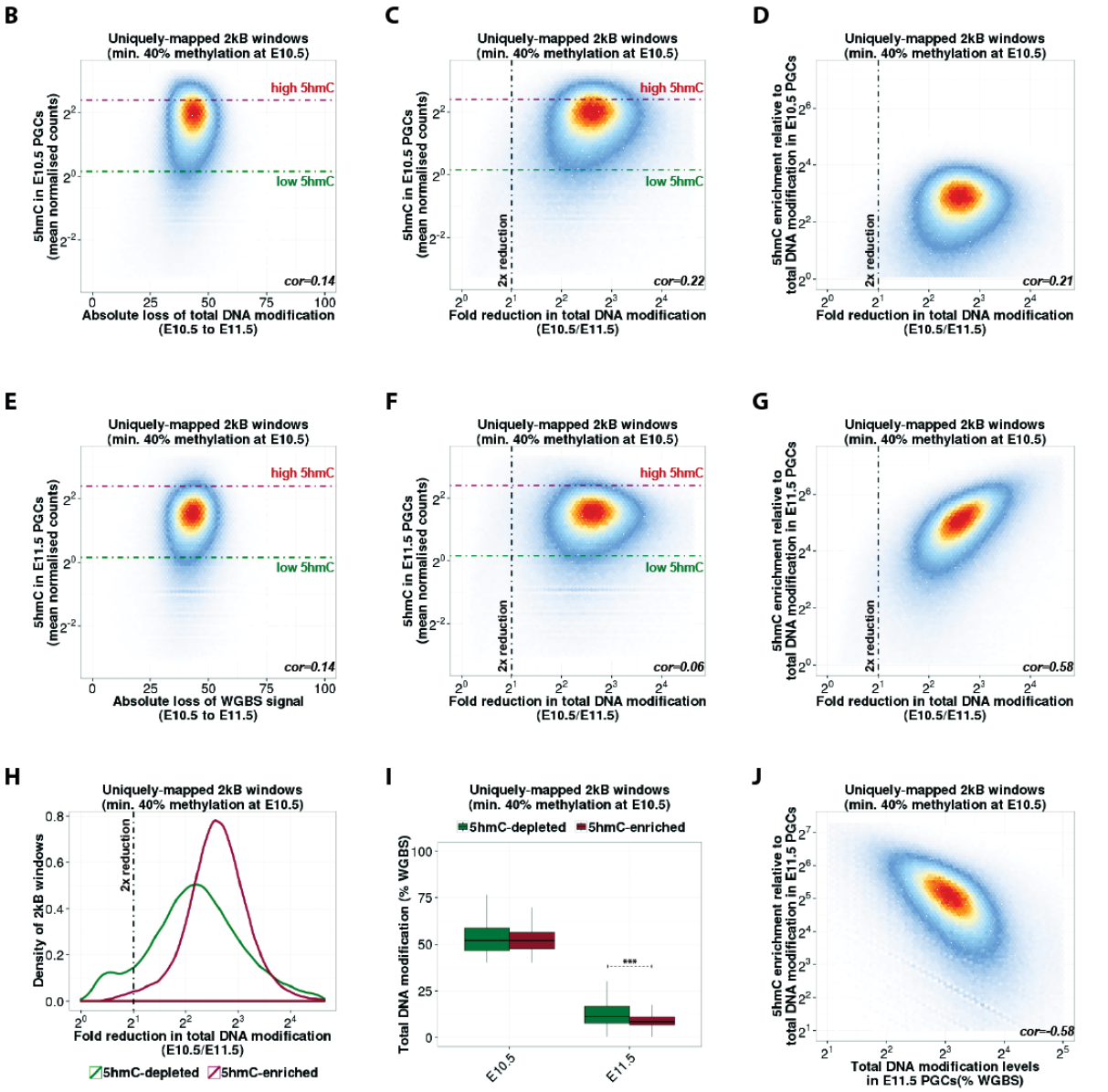


Figure 3.7 –Correlation between 5hmC levels and the extent to which total DNA modification levels decrease between various stages of PGC development. A) *Pairwise Spearman correlation analysis between the decrease in total DNA modification levels between two given stages (either the*

total amount of modification lost or the relative amount of modification lost based on the fold-reduction coefficient) and the 5hmC levels (either the total levels or the levels relative to the total amount of DNA modification) in a given stage for all 2kB windows (with a minimum 40% total DNA modification in E10.5 PGCs). ***B,E)*** *Density heatmap showing correlation between the absolute loss of WGBS signal at all 2kB windows (with minimum 4 CpGs and 40% total DNA modification in E10.5 PGCs) between E10.5 and E11.5 and the total 5hmC level observed at E10.5 (B) or E11.5 (E). Spearman correlation is shown; dashed horizontal red and green lines denotes cutoff for windows significantly enriched or depleted for 5hmC (as determined by adj. Poisson p-value).* ***C,F)*** *Density heatmap showing correlation between the relative decrease in total DNA modification levels (based on the fold-reduction coefficient) at all 2kB windows (with minimum 4 CpGs and 40% total DNA modification in E10.5 PGCs) between E10.5 and E11.5 and the total 5hmC level observed at E10.5 (C) or E11.5 (F). Spearman correlation is shown; dashed horizontal red and green lines denotes cutoff for windows significantly enriched or depleted for 5hmC (as determined by adj. Poisson p-value); dashed vertical black line denotes 2-fold reduction (for reference).* ***D,G)*** *Density heatmap showing correlation between the relative decrease in total DNA modification levels (based on the fold-reduction coefficient) for all 2kB windows (with minimum 4 CpGs and 40% total DNA modification in E10.5 PGCs) between E10.5 and E11.5 and the 5hmC levels relative to the total amount of DNA modification observed at E10.5 (D) or E11.5 (G). Spearman correlation is shown; dashed vertical black line denotes 2-fold reduction (for reference).* ***H)*** *Density plot showing extent of WGBS signal loss for 2kB windows (with minimum 4 CpGs and 40% total DNA modification in E10.5 PGCs) significantly enriched (adj. Poisson p-value < 0.05) for 5hmC at either E10.5 or E11.5 (red) or significantly depleted (adj. Poisson p-value < 0.05) for 5hmC at both E10.5 and E11.5 (green).* ***I)*** *Total DNA modification levels observed at E11.5 for 2kB windows (with minimum 4 CpGs and 40% total DNA modification at E10.5) significantly enriched (adj. Poisson p-value < 0.05) for 5hmC at either E10.5 or E11.5 or significantly depleted (adj. Poisson p-value < 0.05) for 5hmC at both E10.5 and E11.5. p-value based on Mann-Whitney U-test; ***: p-value<0.001.* ***J)*** *Density heatmap showing correlation between the total DNA modification level at all 2kB windows (with minimum 4 CpGs and 40% total DNA modification at E10.5) in E11.5 PGCs and the 5hmC levels relative to the total amount of DNA modification observed at E11.5. Spearman correlation is shown. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.*

3.2.2 Tet1 is not required for global DNA demethylation in gonadal PGCs, but is subsequently needed at a small subset of loci for removal of residual methylation and/or aberrant *de novo* methylation

A model whereby oxidation of 5mC to 5hmC (and its higher oxidative derivatives) maintains a de-methylated state following the major loss of 5mC in PGCs predicts that, upon depletion of the primary 5mC oxygenase, the bulk of 5mC will be lost but some proportion of newly de-methylated loci will

either retain a limited amount of residual methylation or will be targeted for aberrant *de novo* methylation. Previous studies have revealed that Tet1 is the most highly expressed of the Tet family of 5mC oxygenases (at the mRNA level) in wild type PGCs between E10.5 and E13.5 (Hajkova et al., 2010; Yamaguchi et al., 2012). Moreover, published reports have additionally shown that E13.5 female *Tet1*-KO PGCs have no 5hmC as detected by immunofluorescence (Yamaguchi et al., 2012), and that increased total DNA modification levels (as determined by bisulphite sequencing) are observed at certain loci (including some germline promoters and imprint control regions) in both E13.5 male and female *Tet1*-KO germ cells (Yamaguchi et al., 2012; Yamaguchi et al., 2013b). Considering these results, Tet1 presented as the primary 5mC oxygenase in mouse PGCs during gonadal epigenetic reprogramming.

In order to validate my model that 5hmC (and its higher oxidative derivatives) maintains a de-methylated state following the major loss of 5mC, I used a previously published *Tet1*-KO mouse model generated by Jaenisch and colleagues (Dawlaty et al., 2011), which involves deletion of *Tet1* exon 4, resulting in a nonsense mutation. In full *Tet1*-KO mice, deletion of exon 4 from both alleles ultimately leads to no detectable expression of Tet1 protein in any tissue (Dawlaty et al., 2011). I first isolated wild type and *Tet1*-KO PGCs from crosses between *Tet1*-heterozygous males and females. While this knockout model has been extensively validated previously (Dawlaty et al., 2011), I additionally confirmed no detectable expression of exon 4 in isolated *Tet1*-KO PGCs (refer to Chapter 4/Figure 4.2 for details). Moreover, using LC/MS analysis (done in collaboration with Dr Rachel Amouroux), I observed a 93% and 77% reduction in total 5hmC levels respectively in E14.5 female and male *Tet1*-KO PGCs relative to their wild type counterparts (Figure 3.8A). Thus, Tet1 represents the primary 5mC oxygenase in demethylated PGCs, and consequently *Tet1*-KO germ cells represent an appropriate system to test my hypothesis.

I first determined the extent of global DNA demethylation in *Tet1*-KO PGCs by two distinct biochemical methods: 1) I used LC/MS analysis to determine global 5mC levels for *Tet1*-KO and wild type PGCs isolated from embryos at E10.5, E13.5 and E14.5; and 2) I used reduced representation bisulphite sequencing (RRBS) to study locus specific levels of total DNA modification (5mC + 5hmC) at cytosines in a CpG context in PGCs isolated from wild type and *Tet1*-KO E10.5, E12.5 and E14.5 embryos.

RRBS first requires digestion of genomic DNA by MspI, a restriction enzyme which recognizes C^CGG sites, and cuts DNA independent of the methylation state of the internal C; digested DNA is next subject to bisulphite conversion and finally prepared for next generation sequencing (Meissner et al., 2005). While less than 10% of CpGs genome-wide are typically covered by RRBS, those covered are highly enriched for CpG islands, including imprint control regions and germline gene promoters (Bock et al., 2010). Additionally, RRBS has the benefit of being a single-nucleotide resolution mapping method (the advantages of which are discussed above). Between four and ten biological replicates per stage per genotype were used for LC/MS analysis (with DNA from 200-1,000 PGCs per sample used), and two biological replicates per stage were used for the RRBS analysis (with DNA from 1,000-8,000 PGCs per sample used; see Table 3.3 for details). Additionally, for all analyses, male and female PGCs were treated separately from E12.5 onwards to account for any potential sex-specific differences.

Sample	Total aligned reads	CpGs covered (8x)	Mean methylation % (CpG context)	Mean methylation % (non-CpG context)
E10.5 Tet1-KO PGCs (replicate 1)	6,744,305	922,196	12.5%	1.4%
E10.5 Tet1-KO PGCs (replicate 2)	8,173,089	795,951	12.6%	1.8%
E10.5 wild type PGCs (replicate 1)	8,232,875	1,127,784	14.4%	1.5%
E10.5 wild type PGCs (replicate 2)	7,657,619	919,429	27.7%	1.6%
E12.5 female Tet1-KO PGCs (replicate 1)	13,810,865	1,723,053	8.0%	1.7%
E12.5 female Tet1-KO PGCs (replicate 2)	15,730,460	1,775,240	8.7%	1.6%
E12.5 female wild type PGCs (replicate 1)	11,552,013	1,443,707	10.8%	2.0%
E12.5 female wild type PGCs (replicate 2)	12,062,796	1,592,132	9.6%	1.6%
E12.5 male Tet1-KO PGCs (replicate 1)	12,549,636	1,656,192	8.6%	1.7%
E12.5 male Tet1-KO PGCs (replicate 2)	16,467,915	1,751,743	10.9%	1.7%
E12.5 male wild type PGCs (replicate 1)	15,461,491	1,706,350	13.1%	1.5%
E12.5 male wild type PGCs (replicate 2)	15,320,361	1,718,304	12.7%	1.7%
E14.5 female Tet1-KO PGCs (replicate 1)	12,414,703	1,484,733	9.8%	1.7%
E14.5 female Tet1-KO PGCs (replicate 2)	11,543,985	1,400,171	8.9%	2.0%
E14.5 female wild type PGCs (replicate 1)	18,928,983	1,832,650	9.2%	1.6%
E14.5 female wild type PGCs (replicate 2)	11,840,654	1,418,561	9.1%	1.5%
E14.5 male Tet1-KO PGCs (replicate 1)	9,848,480	1,393,878	9.5%	1.8%
E14.5 male Tet1-KO PGCs (replicate 2)	13,850,990	1,518,444	8.5%	1.3%
E14.5 male wild type PGCs (replicate 1)	11,885,257	1,530,631	8.8%	1.5%
E14.5 male wild type PGCs (replicate 2)	11,428,674	1,286,685	9.1%	1.6%

Table 3.3 – Summary of reduced representation bisulphite sequencing (RRBS) statistics

In contrast to what has previously been proposed (Hackett et al., 2013), both LC/MS analysis and RRBS analysis revealed that the bulk of 5mC in PGCs is lost independent of Tet1 (Figures 3.8B-C). Specifically, using RRBS datasets, I first identified all 2kB windows that normally undergo DNA demethylation (minimum 20% total decrease in total DNA modification levels, p-value<0.05) in wild type PGCs between E10.5 and E14.5 (and that are additionally commonly covered in the RRBS datasets derived from PGCs at the earlier and later time points). I next plotted the distribution of total DNA modification levels (as determined by RRBS) in E10.5, E12.5 and E14.5 *Tet1*-KO and wild type PGCs for these regions (Figure 3.8B). This revealed that the major loss in total DNA modification at these regions is observed in wild type PGCs between E10.5 and E12.5, with more limited DNA demethylation observed between E12.5 and E14.5 (Figure 3.8B). Remarkably, very similar DNA demethylation dynamics were observed between *Tet1*-KO and wild type female PGCs (Figure 3.8B). In males on the other hand, E12.5 *Tet1*-KO PGCs appear to have slightly less total DNA modification

than their wild type equivalents, appearing nearly completely hypo-methylated already by this stage. The implication of this particular observation is discussed in detail in Chapter 4.

In support of these initial observations, LC/MS analysis independently confirmed no significant difference in global 5mC levels between *Tet1*-KO and wild type PGCs at E13.5 (male and female) or E14.5 (male), with slightly lower levels of 5mC observed in E14.5 *Tet1*-KO female PGCs relative to wild type (Figure 3.8C). Critically, LC/MS analysis additionally showed that E13.5 and E14.5 PGCs of both genotypes have clearly undergone substantial DNA demethylation when their very low global 5mC levels are compared with the high levels observed in E10.5 PGCs (Figure 3.8C). At the locus specific level, only a limited number of differentially methylated regions are detected in E14.5 *Tet1*-KO PGCs in both males and females, with an equal distribution between hyper- and hypo-methylated regions detected in male PGCs and a slightly higher proportion of hypo-methylated windows detected in female PGCs (Figure 3.8D).

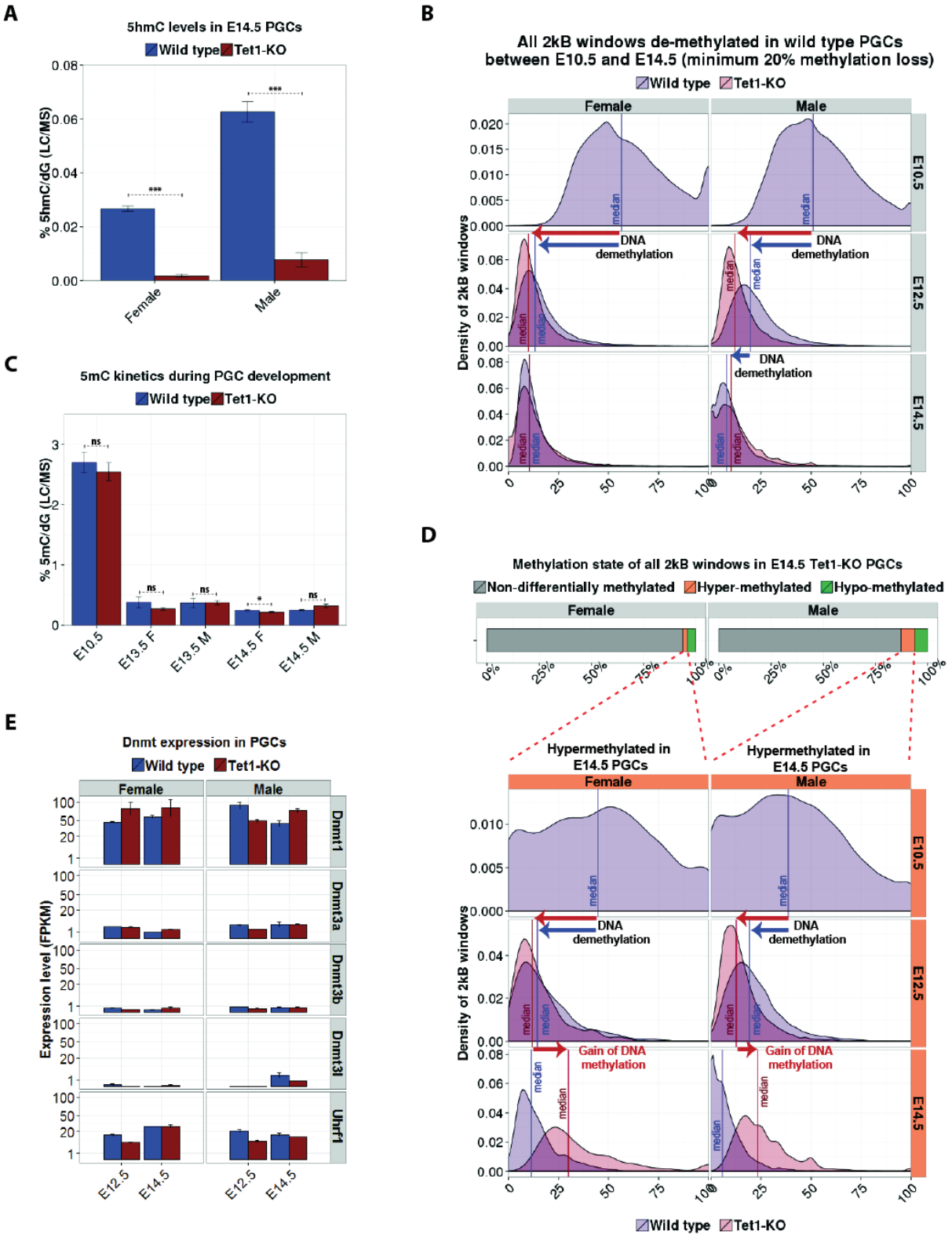


Figure 3.8 – Hyper-methylation in E14.5 Tet1-KO PGCs is associated with a failure to remove residual methylation and/or aberrant de novo methylation in both males and females. A) *Global 5hmC levels in E14.5 wild type and Tet1-KO PGCs. p-values based on Student's t-test. ***: p-value < 0.001.* ***B)*** *Distribution of DNA modification levels in E12.5 and E14.5 female (left) and male (right) Tet1-KO and wild type PGCs at 2kB windows genome-wide with minimum 20% loss of methylation in wild type PGCs between E10.5 and E14.5 (with p-value < 0.05). For comparison, DNA modification*

levels from E10.5 wild type PGCs with same genetic background are also shown. ***C)*** *Global 5mC levels in E10.5, E13.5 and E14.5 wild type and Tet1-KO PGCs. p-values based on Student's t-test. *: p-value < 0.05.* ***D) Top Figure:*** *Proportion of differentially methylated regions (p-value < 0.05, minimum 10% difference in methylation) in Tet1-KO PGCs relative to wild type PGCs isolated from E14.5 female (left) and male (right) embryos.* ***Bottom Figure:*** *Distribution of DNA modification levels in E12.5 (middle panel) and E14.5 (bottom panel) Tet1-KO (red) and wild type (blue) PGCs for E14.5 hyper-methylated 2kB windows (p-value < 0.05, minimum 10% difference in methylation) in females (right) and males (left). For comparison, DNA modification levels from E10.5 wild type PGCs with same genetic background are also shown (top panel). Median total DNA modification levels for each stage and genotype are denoted by vertical lines.* ***E)*** *Expression of factors implicated in DNA methylation in E12.5 and E14.5 PGCs. Note that for all bar charts, bars represent mean value and error bars represent standard error of the mean.*

Conceptually, the few hyper-methylated regions observed in E14.5 *Tet1*-KO PGCs could: 1) represent random regions of hyper-methylation unrelated to gonadal DNA demethylation; 2) represent loci critically dependent on Tet1 for DNA demethylation; or 3) represent regions that have undergone DNA demethylation but either retain a limited amount of residual methylation and/or are targeted for aberrant *de novo* methylation in the absence of Tet1 activity. In order to address these possibilities, I plotted the distribution of total DNA modification levels (as determined by RRBS) in E10.5, E12.5 and E14.5 *Tet1*-KO and wild type PGCs for the few hyper-methylated regions observed in E14.5 *Tet1*-KO germ cells (Figure 3.8D). This analysis revealed that regions hyper-methylated in E14.5 *Tet1*-KO PGCs normally undergo extensive DNA demethylation between E10.5 and E12.5 in wild type PGCs, with more limited DNA demethylation also observed in wild type male PGCs between E12.5 and E14.5 (Figure 3.8D). I reasoned that if Tet1 is critically dependent for DNA demethylation at these regions, significantly increased methylation should also be observed in the absence of Tet1 at E12.5, immediately following the major loss of 5mC in wild type PGCs. Alternatively, if these loci demethylate but either retain a limited amount of residual methylation or are targeted for aberrant *de novo* methylation in the absence of Tet1 activity, total DNA modification levels should be equivalently low at E12.5 with a difference only apparent in the subsequent stage. Indeed, consistent with the latter prediction, both wild type and *Tet1*-KO PGCs are generally hypo-methylated at E12.5 (Figure 3.8D). However, between E12.5 and E14.5, DNA methylation levels increase substantially at these regions in the absence of Tet1, while levels remain stable (female) or undergo a slight further reduction (males) between these stages in wild type PGCs (Figure 3.8D).

In addition to its proposed role as a global DNA demethylase, Tet1 has been attributed a critical function in the demethylation of promoters of certain germline genes, including *Dazl*, and of imprint control regions (Hackett et al., 2013; Yamaguchi et al., 2012; Yamaguchi et al., 2013b). Remarkably, while the *Dazl* promoter and a limited number imprint control regions are indeed hyper-methylated in *Tet1*-KO E14.5 PGCs relative to their wild type equivalents, I observed the same DNA methylation dynamics at these loci as I did more generally. Specifically, I observed no statistically significant difference in total DNA modification levels observed between *Tet1*-KO and wild type PGCs at E12.5 (Figures 3.9A-B). However, in the absence of Tet1, DNA methylation levels increase at these regions between E12.5 and E14.5, while in wild type PGCs, DNA modification levels either remain stable or further decrease to remove the little amount of residual DNA methylation remaining between these two stages (Figures 3.9A-B). Thus, and consistent with the 5hmC dynamics observed in wild type PGCs discussed in detail above, gene knockout studies support a role for Tet1 catalytic activity in maintaining a hypo-methylated state following the major loss of 5mC, rather than in the initial and major phase of DNA demethylation itself.

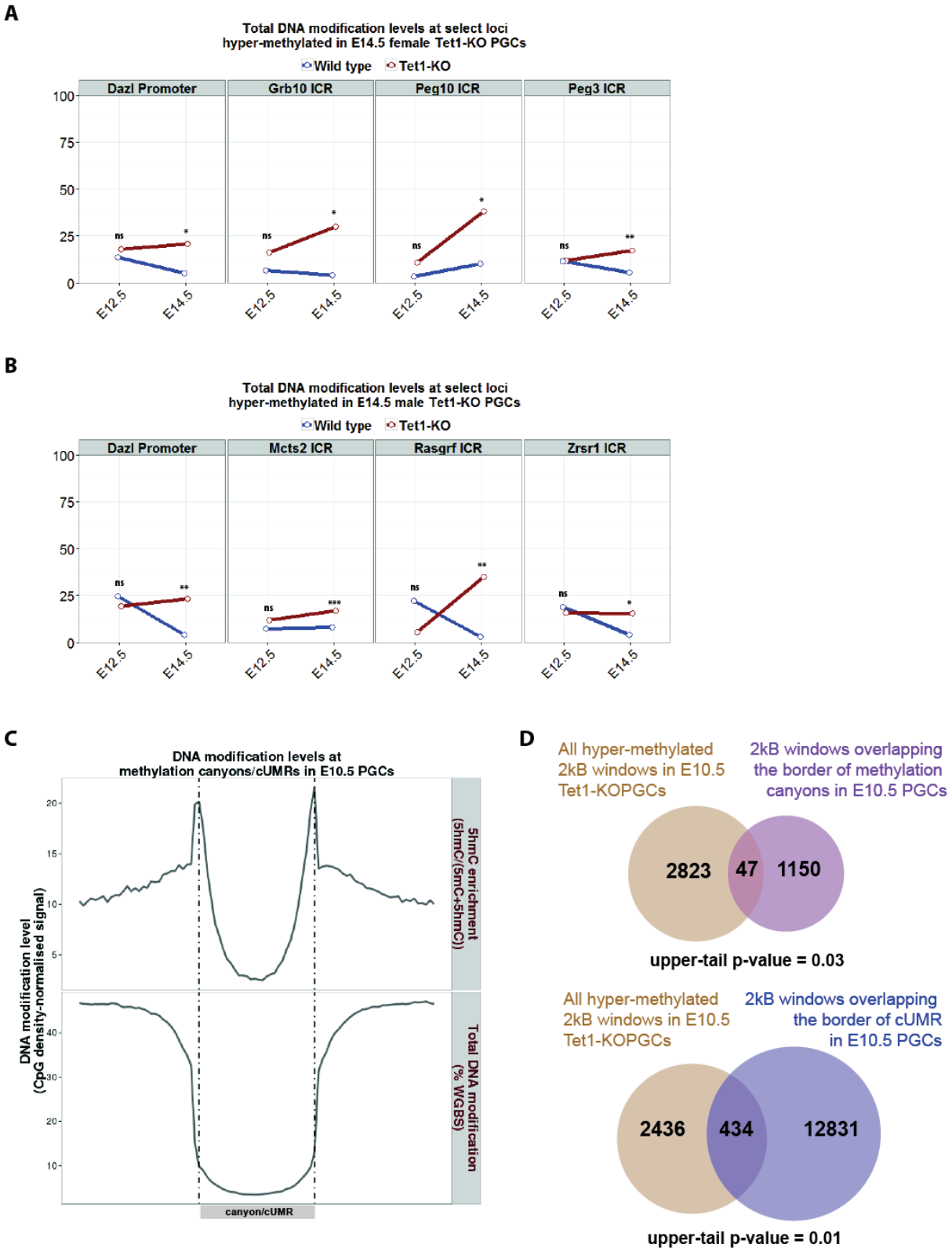


Figure 3.9 – Hyper-methylation in Tet1-KO PGCs is associated with a failure to remove residual methylation and/or aberrant de novo methylation. A-B) *Graphical representation of DNA modification levels in female (A) or male (B) E12.5 and E14.5 Tet1-KO and wild type PGCs for select loci hyper-methylated (p-value < 0.05, minimum 10% difference in methylation) in either female (A) or male (B) E14.5 PGCs. p-values computed by RnBeads software. *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001.* ***C)*** *Metagene plot showing 5hmC levels relative to total DNA modification levels (top panel) and total DNA modification levels (bottom panel) across E10.5 PGC methylation*

canyons and cUMRs. ***D)*** *Overlap between all hyper-methylated 2kB windows in E10.5 Tet1-KO PGCs and 2kB windows overlapping borders of E10.5 methylation canyons (left) or cUMRs (right). p-value based on hyper-geometric test.*

Finally, I considered whether Tet1-mediated 5mC oxidation may have a similar role in maintaining hypo-methylated regions in E10.5 PGCs, prior to global DNA demethylation. In haematopoietic stem cells, 5hmC has been shown to be targeted to the boundaries of large regions of DNA hypo-methylation, termed methylation canyons (>3.5kB) or comparatively unmethylated regions (cUMRs, 1kb-3.5kB), where it appears to help define the borders of these loci by counteracting the activity of the *de novo* methyltransferase machinery (Jeong et al., 2014). With this in mind, I next considered whether 5hmC may act similarly in early PGCs. To do this, I first identified methylation canyons and cUMRs within the methylated E10.5 PGC genome. Consistent with observations in haematopoietic stem cells (Jeong et al., 2014), 5hmC levels appear to be specifically elevated at the borders of these regions in E10.5 PGCs (Figure 3.9C). Moreover, statistically significant enrichment is observed between genomic loci overlapping these borders and the few hyper-methylated regions detected in E10.5 *Tet1*-KO PGCs relative to wild type germ cells at this stage (Figure 3.9D). Thus, one potential role for 5hmC in PGCs prior to gonadal DNA demethylation may be to protect hypo-methylated regions against the aberrant spread of methylation, a model which would be consistent with its aforementioned general role in maintaining a hypo-methylated state in PGCs following the major loss of 5mC.

3.3 Discussion

3.3.1 Gonadal DNA demethylation in PGCs proceeds independent of Tet1

Tet1 has previously been proposed to act as a critical global DNA demethylase during PGC development (Hackett et al., 2013; Yamaguchi et al., 2012; Yamaguchi et al., 2013b). This model was

based on earlier observations that 5hmC levels spike during DNA demethylation at certain imprint control regions (ICRs) and germline gene promoters (Hackett et al., 2013), and that hyper-methylation has been previously observed at these regions in E13.5 *Tet1*-KO PGCs and sperm (Yamaguchi et al., 2012; Yamaguchi et al., 2013b). The results I present in this chapter clearly demonstrate that the bulk of DNA demethylation is not impaired in the absence of Tet1 in mouse primordial germ cells, and thus represents a significant paradigm shift from the current model.

Despite this, I note that my actual observations are in fact consistent with those described previously. Specifically, I too observe that DNA demethylation is accompanied by slightly increased 5hmC levels between E10.5 and E11.5 at ICRs (Figure 3.5D) and germline promoters. However, the apparent increase in 5hmC levels during DNA demethylation predominantly reflects the significantly lower than average levels of 5hmC at most of these regions at E10.5 rather than globally higher than average levels at E11.5 (Figure 3.5D). I also note that LC/MS analysis clearly shows that global levels of 5hmC remain at a constant low level during DNA demethylation in male PGCs, with a slight drop observed following DNA demethylation in female PGCs. These results are in contrast to the reciprocal increase in 5hmC levels predicted by the previously proposed model of global 5mC oxidation followed by passive 5hmC dilution. This discrepancy is likely explained in part by the fact that hMeDIP is not able to accurately compare relative 5hmC levels between samples, while a combined LC/MS and AbaSeq approach can, and also in part by the observation that 5hmC is re-localised from uniquely-mapped to non-uniquely mapped regions (i.e. repetitive elements) between E10.5 and E12.5. Considering this latter point, global 5hmC dynamics as determined by the previously used hMeDIP approach, analysis of which typically involves normalisation only to the uniquely mapped genome, would thus inaccurately suggest a model whereby global 5hmC levels are passively diluted over the course of PGC development.

3.3.2 5hmC mapping in wild type PGCs and methylation dynamics in Tet1-KO PGCs suggest Tet1 acts predominantly to maintain a hypo-methylated state in PGCs

With respect to the hyper-methylation phenotype observed at ICRs and germline gene promoters in *Tet1*-KO PGCs, I too observed slight hyper-methylation at these regions in E14.5 *Tet1*-KO PGCs. I note however that only three of the 19 ICRs investigated fit my definition of differential hyper-methylation (p-value < 0.05, minimum 10% increase in absolute methylation levels) in both female and male germ cells at this stage. Most surprisingly, of the six hyper-methylated ICRs in E14.5 *Tet1*-KO PGCs, five of them increased their methylation levels between E12.5 (at which point they have very low levels of methylation relative to the expected 50% methylation observed in soma) and E14.5, suggesting that hyper-methylation in the absence of Tet1 may be due to other factors beyond impaired DNA demethylation. Indeed, DNA demethylation generally appears to proceed independent of Tet1, and DNA methylation kinetics for hyper-methylated regions in E14.5 *Tet1*-KO PGCs are most consistent with a model of aberrant *de novo* methylation following DNA demethylation, rather than impaired global DNA demethylation itself. At this point it should be noted that I was clearly able to detect low levels of expression of *Dnmt3a* and *Dnmt3b* in PGCs at E12.5 and E14.5 (Figure 3.8E); in contrast, no detectable expression of *Dnmt3l* was observed, suggesting that *de novo* methylation in PGCs between these stages occurs independent of this co-factor (Figure 3.8E). Considering the results I describe in this chapter, the role of the *de novo* methyltransferase activity during gonadal DNA demethylation in PGCs represents an exciting area of future investigation. However, at this point, the exact function of low levels of *de novo* DNA methyltransferase activity in gonadal PGCs, as well as the overall biological significance of the apparent two-tier demethylation process observed, remains to be established.

Altogether, the observations presented above suggest a model whereby Tet1 activity is important to maintain a hypo-methylated state following the major loss of 5mC in PGCs (Figure 3.10). In this

context, a similar role for Tet1 in maintaining regions of hypo-methylation have been recently described in both differentiated somatic cells (Jin et al., 2014) and mouse embryonic stem cells (Williams et al., 2011), as well as for Tet3 in the zygote following the global wave of DNA demethylation observed immediately after fertilisation (Amouroux et al., 2016). Thus, protecting the hypo-methylated state appears to represent a highly conserved role for Tet proteins in mammals.

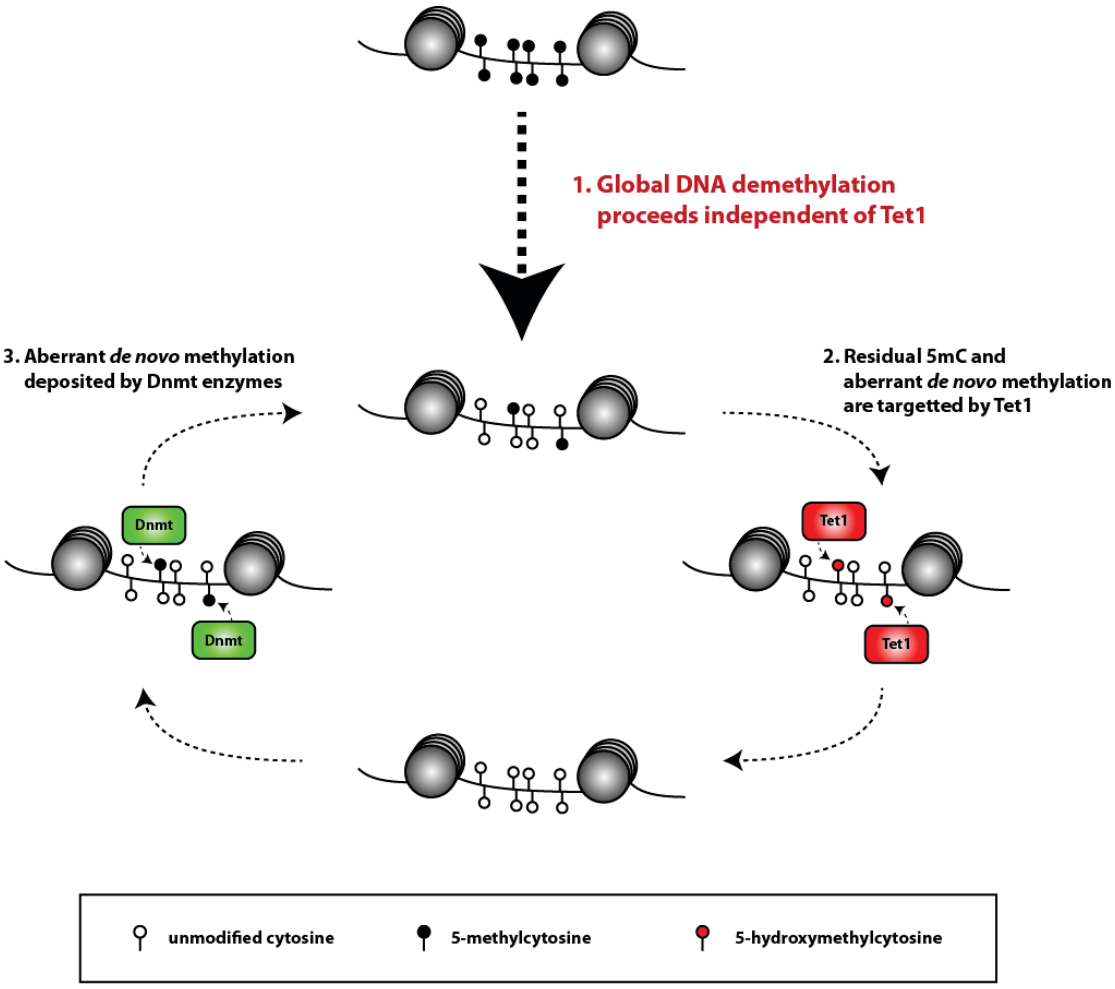


Figure 3.10 –Model. *Tet1 activity helps protect the hypo-methylated state in PGCs following the major loss of 5mC.*

Chapter 4

Results

On the mechanism of Tet1 transcriptional regulation in the mouse foetal germ line

4.1 Introduction

4.1.1 Tet1 regulates expression of meiosis genes in mouse PGCs

Soon after the discovery of Tet1 enzymatic activity, Zhang and colleagues identified Tet1 as a key regulator of female germ cell development (Yamaguchi et al., 2012). Specifically, to explore a potential role of Tet1 in germ-cell development, the authors first generated Tet1 gene-trap mice (Yamaguchi et al., 2012). Tet1 protein normally contains Cxxc domain towards the amino terminus, which binds preferentially to unmethylated CpG-rich DNA (Williams et al., 2011), and a carboxy-terminal catalytic core region that includes a Cys-rich insert and a larger double-stranded β-helix (DSBH) domain (Figure 4.1) (Pastor et al., 2013). In the homozygous Tet1 gene-trap mice generated by Zhang and colleagues, a single site insertion of a tandem-repeated gene-trap cassette into the first intron of the Tet1 gene resulted in expression of a fusion protein between the first exon of Tet1, which importantly still contains a functional Cxxc DNA binding domain, and the β-galactosidase and neomycin resistance domains (Yamaguchi et al., 2012).

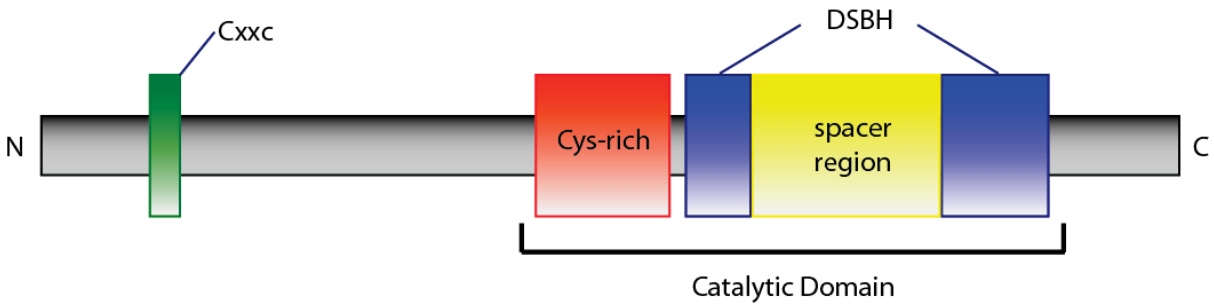


Figure 4.1 – Domain organisation of Tet1 protein.

Using this gene-trap model, the authors found that *Tet1*-KO ovaries are significantly smaller than their wild type counterparts, with asymmetric ovary size caused by ovarian agenesis frequently observed in the knockout animals (Yamaguchi et al., 2012); additionally, they observed a significant reduction in

oocyte number starting from E16.5, concurrent with a significant increase in the observed number of apoptotic oocytes (Yamaguchi et al., 2012). At the molecular level, using immunostaining against the meiotic marker Sycp3 and the centromere-marker CREST, the authors show that about 50% of oocytes remained at the leptotene stage and no oocytes reached the pachytene stage in E16.5 *Tet1*-KO ovaries (Yamaguchi et al., 2012). Moreover, they observed obvious defects in synapsis in zygotene stage oocytes and a clear accumulation of γH2AX in pachytene- and early diplotene-stage oocytes (Yamaguchi et al., 2012). Further RNA-Seq analysis on E13.5 female *Tet1*-KO PGCs, which normally would have just initiated the meiosis programme, revealed significantly reduced expression of meiosis-specific genes (Yamaguchi et al., 2012).

To investigate how Tet1 might be involved in activation of meiotic genes, Zhang and colleagues performed whole-genome bisulphite sequencing (WGBS) analysis. It should be noted, however, that the authors only obtained 2x genome coverage which does not allow robust identification of differentially methylated regions, and hence it is difficult to draw meaningful conclusions from this dataset. Nonetheless, Zhang and colleagues also performed traditional high-coverage targeted bisulphite sequencing of the promoters of three verified Tet1-target genes (*Sycp1*, *Mael* and *Sycp3*) and observed increased methylation levels at these loci in E13.5 *Tet1*-KO PGCs (Yamaguchi et al., 2012). Based on these results, the authors suggested that Tet1-mediated DNA demethylation of meiosis genes is probably involved in their activation during PGC development (Yamaguchi et al., 2012). It should be noted, however, that some downregulated genes, such as *Stra8*, showed no obvious change in DNA methylation indicating that they are either regulated indirectly by Tet1 or regulated in a DNA-methylation-independent manner (Yamaguchi et al., 2012). Considering this and the results I present in Chapter 3, which clearly show that Tet1 does not act as a global DNA demethylase during gonadal epigenetic reprogramming, I set out to further characterise the nature of the transcriptional defect in foetal germ cells devoid of Tet1.

4.1.2 Potential mechanisms of Tet1 transcriptional regulation

Tet1 was recognized early on as a transcription regulator, capable of modulating gene expression by a variety of mechanisms. It was originally hypothesized that Tet1 may regulate transcription by inducing locus specific DNA demethylation at regions where it is recruited (Figure 4.2A) (Costa et al., 2013; Gao et al., 2013; Yamaguchi et al., 2012). For example, Tet1 was proposed to demethylate regulatory elements of pluripotency specific genes, such as *Oct4* (Gao et al., 2013), during iPS reprogramming, a process which involves the ectopic expression of pluripotency transcription factors in somatic cells, reprogramming them to an embryonic stem cell-like state (Takahashi and Yamanaka, 2006). In this context, Tet1-mediated DNA demethylation of *Oct4* regulatory elements was hypothesized to drive transcriptional activation of the *Oct4* gene (Gao et al., 2013). It should be noted, however, that subsequent follow-up studies have been unable to validate these initial findings (Hu et al., 2014). Nonetheless, inducing locus-specific DNA demethylation presents an intuitive potential mechanism by which Tet1 may regulate transcription (Figure 4.2A).

Tet1 enzymatic activity has also been shown to regulate transcription beyond its proposed role in DNA demethylation. Specifically, using *Tdg*-knockout embryonic stem cells, which allows for the accumulation of 5fC and 5caC, He and colleagues showed that recruitment of the enhancer-associated histone acetyltransferase p300 correlates strongly with Tet1 occupancy and is recruited via binding to these oxidised 5mC derivatives (Figure 4.2B) (Song et al., 2013a). Additionally, 5hmC, 5fC, and 5caC are all enriched at enhancer elements in wild type mouse ESCs (Song et al., 2013a; Yu et al., 2012), and have recently been shown to be stable modifications in mammalian DNA in their own right (Bachman et al., 2015; Bachman et al., 2014). Considering these observations and that p300 acetylates H3K27, which is generally associated with enhancer activation (Creyghton et al., 2010), oxidation of 5mC at enhancers may be an important mechanism by which Tet1 regulates transcription (Figure 4.2B). It should also be noted that Tet proteins appear to play an important role in maintaining

a hypo-methylated state at enhancer elements (Hon et al., 2014; Lu et al., 2014), although this appears to be predominantly driven by Tet2 rather than Tet1 in mouse ESCs (Hon et al., 2014). Additionally, *in vitro* biochemical studies have recently shown that the presence of 5fC and 5caC in the DNA template shifts RNA polymerase II (Pol II) from an active population (poised for elongation) to a paused population, and thus promotes Pol II stalling and RNA backtracking (Figure 4.2C) (Kellinger et al., 2012). Further structural analysis revealed that Pol II is directly capable of sensing the DNA oxidative methylation state through its conserved epi-DNA recognition loop (Wang et al., 2015). Moreover, analysis of transcription elongation rates in mouse embryonic stem cells (mESCs) further confirmed that Pol II transiently slows down specifically at 5fC/5caC sites during transcription *in vivo* (Wang et al., 2015).

In addition to its enzymatic function, Tet1 can regulate transcription via non-enzymatic mechanisms involving recruitment of diverse chromatin modifiers to Tet1-binding sites (Figure 4.2D). In a seminal study, Helin and colleagues identified a direct interaction between Tet1 and the histone deacetylase complex component Sin3a using a tandem purification mass spectrometry approach (Williams et al., 2011). Additionally, using ChIP studies, the authors noted that Sin3a has a similar binding profile as Tet1 in mESCs, and that loss of Tet1 results in significantly reduced recruitment of Sin3a (Williams et al., 2011). As Tet1 depletion in mESCs resulted in up-regulation of a large subset of genes, the authors concluded that Tet1 has a role in transcriptional repression, in part mediated via recruitment of Sin3a deacetylase activity (Figure 4.2D) (Williams et al., 2011). Remarkably, the authors of this study additionally showed that the transcriptional effects observed following Tet1 knockdown in wild type and *Dnmt1*/*Dnmt3a*/*Dnmt3b*-triple knockout mESCs are similar, indicating that these transcriptional effects are predominantly independent of loss of Tet1 catalytic activity (Williams et al., 2011). The study by Helin and colleagues was expanded upon by Pasini and colleagues, who performed mass spectrometry on proteins interacting with tagged O-linked N-acetylglucosamine transferase (Ogt), and identified both Tet1 and Tet2, as well as a number of components of the Sin3a complex (Vella et al., 2013). Intriguingly, and very similar to what was described by Helin and

colleagues, the authors of this study observed significant overlap between Tet1 and Ogt binding sites, and, upon Tet1 depletion, reduced recruitment of Ogt to chromatin concurrent with general up-regulation genes associated Tet1/Ogt binding, suggesting that direct recruitment of Ogt to promoters could inhibit transcription (Figure 4.2D) (Vella et al., 2013). Intriguingly, Ogt binding to Tet1 appears to induce its O-GlcNAcylation, and mutation of the putative O-GlcNAcylation site on Tet1 leads to decreased level of Tet1 protein, thus hinting at a complex regulatory relationship between Tet1 and Ogt (Shi et al., 2013; Vella et al., 2013). Tet1 has also been shown to directly interact with the Ezh2 and Suz12 components of polycomb repressive complex 2 (PRC2), which is responsible for deposition of tri-methylation at H3K27 (Neri et al., 2013). It should be noted, however, that other studies have been unable to corroborate these findings ((Wu et al., 2011); R. Amouroux, personal communication). Nonetheless, Tet1 loss is clearly associated with reduced recruitment of Ezh2 to chromatin in mouse ESCs, concurrent with hyper-methylation of DNA at these regions (Wu et al., 2011). However, whether the recruitment is direct or indirect and what the functional importance of the Tet1-PRC2 interaction on gene expression is remain unclear.

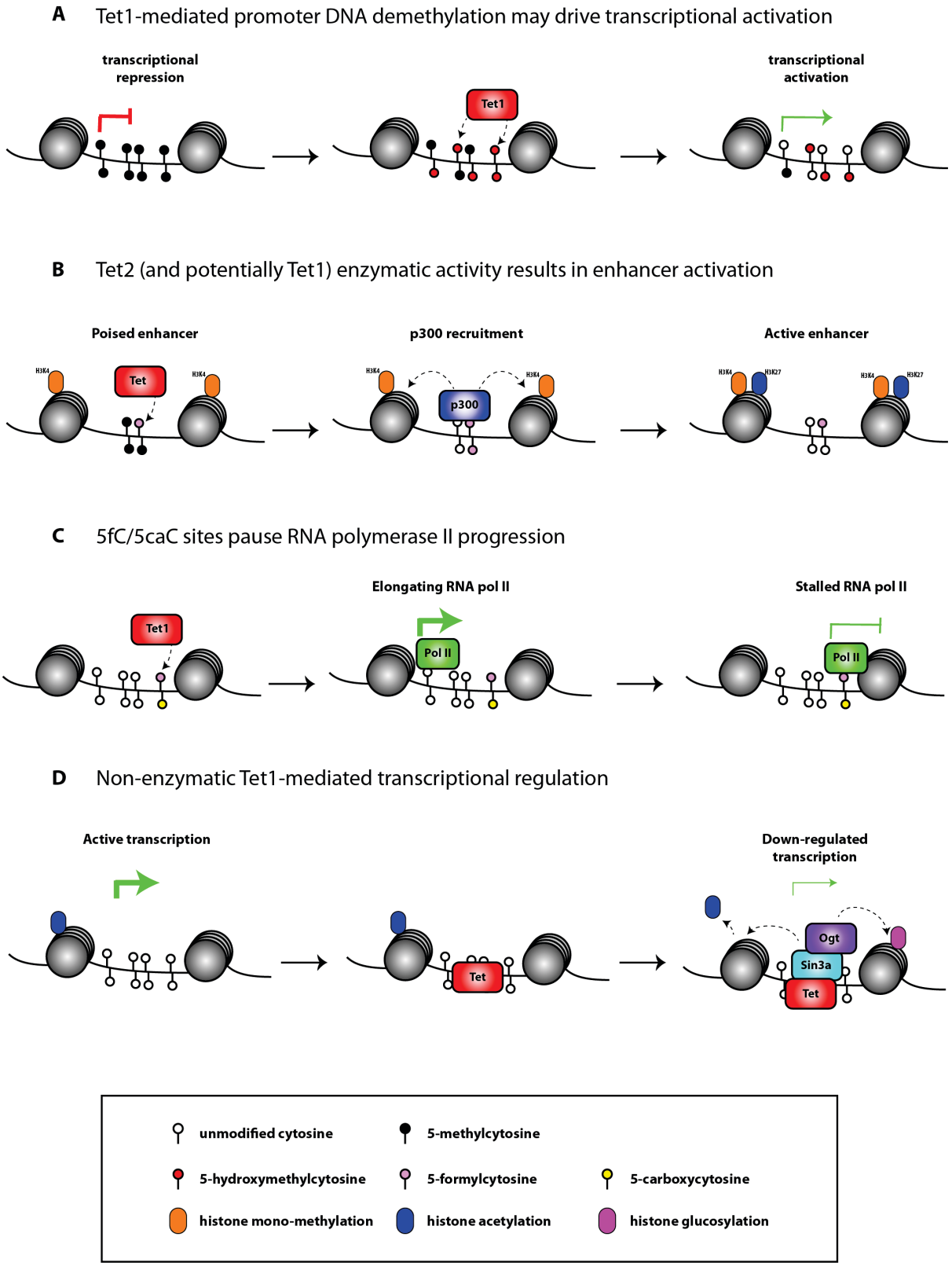


Figure 4.2 – Known mechanisms of transcriptional regulation by Tet1 – adapted from (Hill et al., 2014). ***A)*** *Tet1-mediated promoter DNA demethylation may drive transcriptional activation of target genes.* ***B)*** *5mC oxidation of enhancer regions results in increased p300 recruitment and consequently enhancer activation.* ***C)*** *The presence of 5fC/5caC is known to stall RNA polymerase II (Pol II) elongation rates, and thus reduce the total amount of mRNA transcribed from a locus.* ***D)*** *Direct or indirect recruitment of chromatin modifiers to Tet1 binding sites.*

4.2 Results

4.2.1 Loss of Tet1 results in increased stability of the demethylated pre-meiotic germ cell state

To understand how Tet1 regulates entry into meiosis, I first generated twelve RNA-Seq datasets derived from E10.5 (male/female), E12.5 female and E14.5 female *Tet1*-KO (Dawlaty et al., 2011) and wild type PGCs to study the dynamics of meiotic gene expression. Additionally, I carried out transcriptional profiling from PGCs isolated from E12.5 and E14.5 male *Tet1*-KO (Dawlaty et al., 2011) and wild type embryos for direct comparison, as male germ cells do not enter into meiosis until after birth. Of note, each RNA-Seq biological replicate represented a pool of RNA from between two to six embryos to control for embryo-to-embryo variability; moreover, both wild type and *Tet1*-KO PGCs were isolated from embryonic littermates of *Tet1*-heterozygous (Dawlaty et al., 2011) x *Tet1*-heterozygous (Dawlaty et al., 2011) matings to carefully control for background effects and developmental staging. The complete dataset consisted of two biological replicates each of both wild type and *Tet1*-KO foetal germ cells at five distinct stages of development - 20 samples in total (see Table 4.1 for details).

I first determined how expression of *Tet1* and its homologs *Tet2* and *Tet3* is affected in PGCs depleted of Tet1. Consistent with the predicted expression patterns in this particular knockout model (Dawlaty et al., 2011), detectable expression of *Tet1* mRNA is observed at all tested stages, although its levels are significantly reduced in both E12.5 male and female *Tet1*-KO PGCs (Figures 4.3C-G). Importantly, however, all *Tet1*-KO PGCs showed no detectable expression of the deleted exon 4 (Figures 4.3A, 4.3C-G): in this knockout model, deletion of exon 4 results in a nonsense mutation, yielding an unstable truncated product lacking the catalytic domain of Tet1 (Dawlaty et al., 2011). Thus, although residual Tet1 transcript can be detected in this model (Figures 4.3C-G and (Dawlaty et

al., 2011)), no full-length transcript or truncated Tet1 protein was observed in either PGCs or soma of *Tet1*-KO embryos (Figures 4.3B-G and (Dawlaty et al., 2011)).

Consistent with these observations, *Tet1*-KO PGCs isolated from embryos at all stages showed significant reduction of total 5hmC levels (as computed by LC/MS analysis) compared with those observed in their wild type counterparts (Figure 4.3G). Intriguingly, the relative decrease in 5hmC levels increases with developmental stage, with E10.5 *Tet1*-KO PGCs showing an approximately 50% reduction, while E14.5 *Tet1*-KO PGCs show approximately 85% loss of 5hmC (Figure 4.3G). This suggests that foetal germ cells are increasingly dependent on Tet1 for 5mC oxidation as they progress through their development (Figure 4.3H). Surprisingly, neither transcriptional up-regulation of *Tet2* nor *Tet3* appears to compensate for Tet1 loss during foetal germ cell development between E10.5 and E14.5 (Figures 4.3C-H). Rather, *Tet2* and *Tet3* expression in *Tet1*-KO PGCs at most stages is in line with the expression levels observed in their wild type counterparts, with *Tet3* expression significantly reduced in E12.5 female *Tet1*-KO PGCs (Figures 4.3C-G).

Sample	Total aligned paired end reads	Total aligned single end reads	Total aligned reads	Number of genes detected (FPKM > 0.1)
E10.5 Tet1-KO PGCs (replicate 1)	27,934,274	6,531,874	34,466,148	13,690
E10.5 Tet1-KO PGCs (replicate 2)	34,681,316	6,854,408	41,535,724	13,786
E10.5 wild type PGCs (replicate 1)	31,962,294	6,739,387	38,701,681	15,322
E10.5 wild type PGCs (replicate 2)	40,025,800	4,739,626	44,765,426	15,237
E12.5 female Tet1-KO PGCs (replicate 1)	44,855,472	11,364,324	56,219,796	15,065
E12.5 female Tet1-KO PGCs (replicate 2)	26,457,358	6,927,142	33,384,500	14,824
E12.5 female wild type PGCs (replicate 1)	29,401,150	7,550,809	36,951,959	16,156
E12.5 female wild type PGCs (replicate 2)	36,491,284	7,380,861	43,872,145	15,323
E12.5 male Tet1-KO PGCs (replicate 1)	23,402,232	5,058,571	28,460,803	14,589
E12.5 male Tet1-KO PGCs (replicate 2)	28,895,882	6,795,491	35,691,373	14,843
E12.5 male wild type PGCs (replicate 1)	27,657,008	6,346,269	34,003,277	15,697
E12.5 male wild type PGCs (replicate 2)	28,789,312	6,125,174	34,914,486	15,998
E14.5 female Tet1-KO PGCs (replicate 1)	25,438,028	5,998,085	31,436,113	14,996
E14.5 female Tet1-KO PGCs (replicate 2)	38,611,124	7,569,699	46,180,823	13,796
E14.5 female wild type PGCs (replicate 1)	38,762,174	8,460,970	47,223,144	15,269
E14.5 female wild type PGCs (replicate 2)	17,060,540	4,402,909	21,463,449	15,293
E14.5 male Tet1-KO PGCs (replicate 1)	30,218,364	6,753,119	36,971,483	15,200
E14.5 male Tet1-KO PGCs (replicate 2)	33,295,938	7,293,839	40,589,777	15,043
E14.5 male wild type PGCs (replicate 1)	32,629,876	6,775,115	39,404,991	14,480
E14.5 male wild type PGCs (replicate 2)	47,552,152	10,785,382	58,337,534	15,332

Table 4.1 – Summary of RNA-Seq sequencing statistics

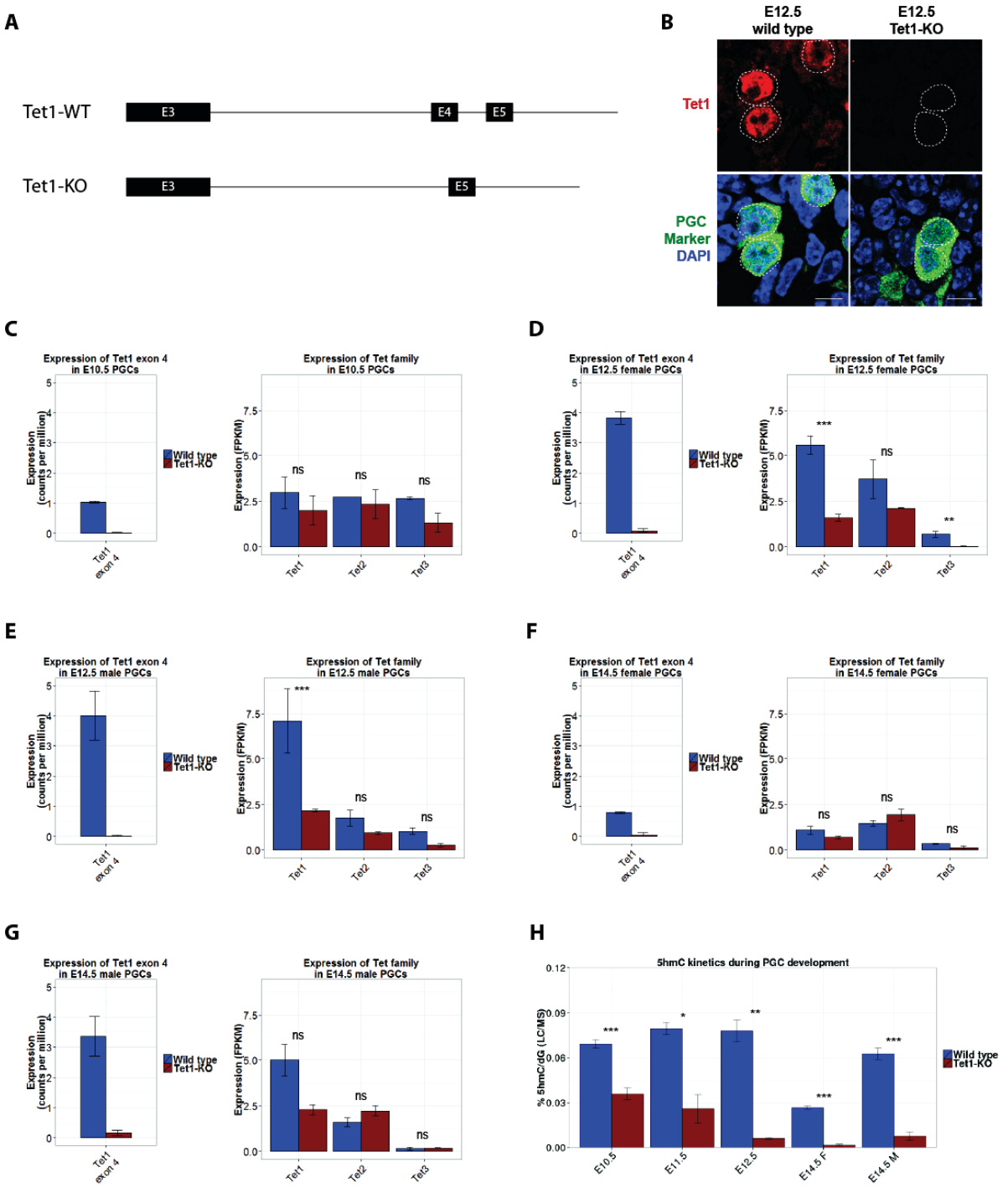


Figure 4.3–Tet protein expression and 5hmC levels in Tet1-KO PGCs. A) *Schematic of Tet1 exon 3-5 in wild type and Tet1-KO cells. Adapted from (Dawlaty et al., 2011).* ***B)*** *Immunostaining for N-terminus of Tet1 protein in E12.5 wild type and Tet1-KO PGCs. PGCs are marked in green with an anti-GFP antibody.* ***C-G)*** *Expression level of the deleted Tet1 exon 4 (left) and of Tet1, Tet2 and Tet3 full length transcripts in E10.5 (C), E12.5 female (D), E12.5 male (E), E14.5 female (F), and E14.5 male (G) Tet1-KO and wild type PGCs. adjusted p-value (adj. p-value) computed by DESeq2 is shown; *: adj. p-value < 0.05, **: adj. p-value < 0.01, *** adj. p-value < 0.001.* ***H)*** *Total 5hmC levels in E10.5 to E14.5 Tet1-KO and wild type PGCs. p-values based on Student's t-test; *: p-value*

*< 0.05, **: p-value < 0.01, *** p-value < 0.001. In all cases, bars represent mean value and error bars represent standard error of the mean.*

With respect to the general transcriptional trends observed in *Tet1*-KO foetal germ cells, initial data analysis revealed a surprisingly complex relationship between Tet1 loss and the transcriptional progression of PGCs. On this (predominantly C57B6) genetic background, significant transcriptional changes are apparent between E10.5 and E12.5 in both wild type male and female germ cells, with sex specific differences already noticeable (Figure 4.3A). Of note, between E12.5 and E14.5, the developmental profile of wild type female germ cells diverges considerably from the initial transcriptional trend observed between E10.5 and E12.5, as female germ cells enter into the meiotic programme at this time (Figure 4.3A). On the other hand, between E12.5 and E14.5, male germ cells appear to continue along the same the transcriptional pathway as initially observed between E10.5 and E12.5 (Figure 4.3A). In the absence of Tet1, principle component analysis revealed that PGCs progress along the male and female developmental pathways with remarkably different kinetics than those just described. E10.5 *Tet1*-KO PGCs appear to have progressed further than their wild type counterparts along principle component (PC) 1, moving towards the positions occupied by E12.5 and E14.5 PGC samples (Figure 4.3A). Likewise, E12.5 male and female *Tet1*-KO PGCs also appear to have progressed further along PC1 than their wild type equivalents (Figure 4.4A); moreover, these are both displaced downwards along PC2, towards the position specifically occupied by E14.5 male PGCs (Figure 4.4A). In contrast to the earlier (E10.5 and E12.5) PGCs, E14.5 *Tet1*-KO foetal germ cells are shifted towards the left along PC1, suggesting that their developmental progression may be delayed at this time-point (Figure 4.4A). While this trend is observed in both males and females, it is considerably stronger in female PGCs (Figure 4.4A). In addition, E14.5 *Tet1*-KO female germ cells are clearly displaced downwards along PC2, towards the positions occupied by E12.5 and E14.5 male PGCs. In males, in contrast, a very slight shift upwards is observed, nudging E14.5 *Tet1*-KO male PGCs ever so slightly towards the position occupied by E12.5 male samples (Figure 4.4A).

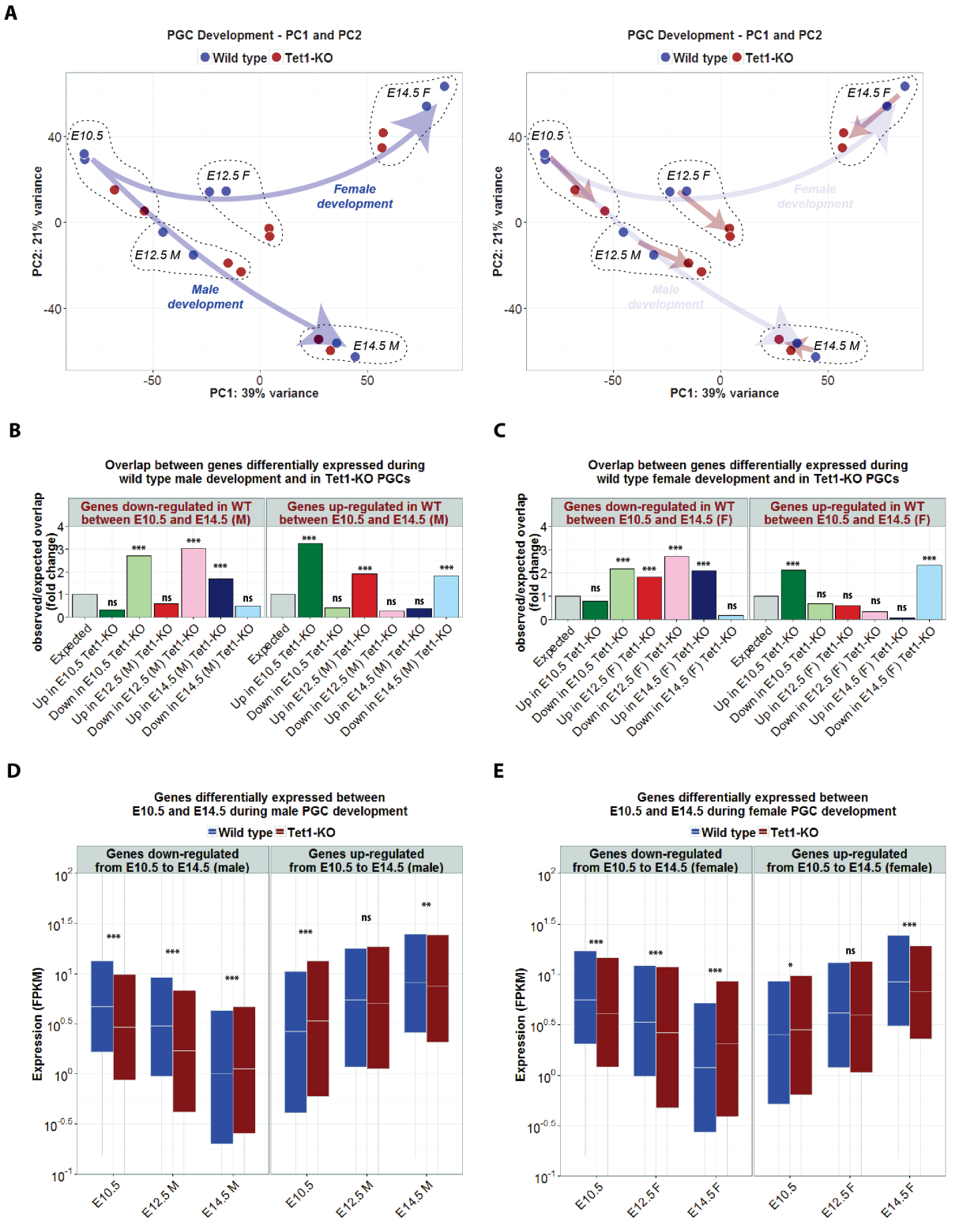


Figure 4.4 –Loss of Tet1 promotes increased stability of the demethylated pre-meiotic germ cell state. A) *Principle component analysis showing clustering of RNA-Seq samples derived from Tet1-KO and wild type E10.5, E12.5 (male and female) and E14.5 (male and female) PGCs. Blue arrows on the left show normal wild type developmental progression for female and male PGCs between E10.5 and E14.5; red arrows on the right show displacement of Tet1-KO PGCs along principle component (PC) 1 and PC2 relative to their wild type sex- and stage-matched equivalents.* ***B)*** *Overlap between*

genes differentially up- or down-regulated in Tet1-KO E10.5, E12.5 male and E14.5 male PGCs and genes differentially down-regulated (left panel) or up-regulated (right panel) between E10.5 and E14.5 in wild type male PGCs p-value computed by hyper-geometric test – see Table 4.2 for details. ***C)*** *Overlap between genes differentially up- or down-regulated in Tet1-KO E10.5, E12.5 female and E14.5 female PGCs and genes differentially down-regulated (left panel) or up-regulated (right panel) between E10.5 and E14.5 in wild type female PGCs. p-value computed by hyper-geometric test – see Table 4.2 for details.* ***D)*** *Distribution of gene expression levels in E10.5, E12.5 and E14.5 male Tet1-KO and wild type PGCs for all genes differentially down-regulated (left panel) or up-regulated (right panel) between E10.5 and E14.5 in wild type male PGCs. p-values based on Mann-Whitney U-test *: p-value < 0.05, **: p-value < 0.01, *** p-value < 0.001.* ***E)*** *Distribution of gene expression levels in E10.5, E12.5 and E14.5 female Tet1-KO and wild type PGCs for all genes differentially down-regulated (left panel) or up-regulated (right panel) between E10.5 and E14.5 in wild type female PGCs. p-values based on Mann-Whitney U-test *: p-value < 0.05, **: p-value < 0.01, *** p-value < 0.001. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.*

To complement this initial analysis, I next identified genes differentially up- and down-regulated during wild type male and female development between E10.5 and E14.5 (adjusted p-value < 0.05, DESeq2), and studied how the expression kinetics for these genes are affected in the absence of Tet1 (Figures 4.4B-E). Consistent with principle component analysis, genes differentially up-regulated in E10.5 *Tet1*-KO PGCs are significantly enriched for genes up-regulated during wild type male and female development, while genes differentially down-regulated in E10.5 *Tet1*-KO PGCs are significantly enriched for genes down-regulated during the development of both sexes (Figures 4.4B-C, Table 4.2). Moreover, generally genes developmentally up- and down-regulated in wild type male (Figure 4.4D) and female (Figure 4.4E) germ cells are respectively higher and lower expressed in E10.5 *Tet1*-KO PGCs. Of note, I observed a more significant increase among genes up-regulated during male development in E10.5 PGCs lacking Tet1 (Figure 4.4D-E, Table 4.2).

E14.5 *Tet1*-KO PGCs show the opposite relationship with respect to developmentally regulated gene sets. Specifically, genes differentially up-regulated in E14.5 male and female *Tet1*-KO PGCs are significantly enriched for genes developmentally down-regulated in wild type male (Figure 4.4B, Table 4.2) and female (Figure 4.4C, Table 4.2) germ cells, respectively. Furthermore, genes differentially down-regulated in E14.5 male and female *Tet1*-KO PGCs are significantly enriched for

genes up-regulated during wild type male (Figure 4.4B, Table 4.2) and female (Figure 4.4C, Table 4.2) development, respectively. Similarly, genes developmentally up-regulated in wild type male and female germ cells are generally significantly lower expressed in E14.5 male (Figure 4.4D) and female (Figure 4.4E) *Tet1*-KO germ cells, respectively, while the reciprocal relationship is observed for genes down-regulated during the development of each sex (Figures 4.4D-E).

Group 1	Group 2	Overlap	Genes in group 1	All genes	Genes in group 2	p-value	Relative enrichment
Genes up-regulated in WT between E10.5 and E14.5 (M)	Genes up-regulated in Tet1-KO PGCs at E10.5	637	1070	17120	3148	1.95E-214	3.24
Genes up-regulated in WT between E10.5 and E14.5 (M)	Genes down-regulated in Tet1-KO PGCs at E10.5	130	1695	17120	3148	1	0.42
Genes up-regulated in WT between E10.5 and E14.5 (M)	Genes up-regulated in Tet1-KO PGCs at E12.5 (M)	79	191	14580	3148	1.78E-10	1.92
Genes up-regulated in WT between E10.5 and E14.5 (M)	Genes down-regulated in Tet1-KO PGCs at E12.5 (M)	44	773	14580	3148	1	0.26
Genes up-regulated in WT between E10.5 and E14.5 (M)	Genes up-regulated in Tet1-KO PGCs at E14.5 (M)	30	261	10770	3148	1	0.39
Genes up-regulated in WT between E10.5 and E14.5 (M)	Genes down-regulated in Tet1-KO PGCs at E14.5 (M)	60	113	10770	3148	2.80E-08	1.82
Genes down-regulated in WT between E10.5 and E14.5 (M)	Genes up-regulated in Tet1-KO PGCs at E10.5	72	1070	17120	3420	1	0.34
Genes down-regulated in WT between E10.5 and E14.5 (M)	Genes down-regulated in Tet1-KO PGCs at E10.5	911	1695	17120	3420	2.39E-238	2.69
Genes down-regulated in WT between E10.5 and E14.5 (M)	Genes up-regulated in Tet1-KO PGCs at E12.5 (M)	27	191	14580	3420	1	0.60
Genes down-regulated in WT between E10.5 and E14.5 (M)	Genes down-regulated in Tet1-KO PGCs at E12.5 (M)	546	773	14580	3420	4.94E-182	3.01
Genes down-regulated in WT between E10.5 and E14.5 (M)	Genes up-regulated in Tet1-KO PGCs at E14.5 (M)	139	261	10770	3420	1.11E-13	1.68
Genes down-regulated in WT between E10.5 and E14.5 (M)	Genes down-regulated in Tet1-KO PGCs at E14.5 (M)	18	113	10770	3420	1	0.50
Genes up-regulated in WT between E10.5 and E14.5 (F)	Genes up-regulated in Tet1-KO PGCs at E10.5	478	1070	17120	3609	1.49E-72	2.12
Genes up-regulated in WT between E10.5 and E14.5 (F)	Genes down-regulated in Tet1-KO PGCs at E10.5	247	1695	17120	3609	1	0.69
Genes up-regulated in WT between E10.5 and E14.5 (F)	Genes up-regulated in Tet1-KO PGCs at E12.5 (F)	29	192	14580	3609	1	0.61
Genes up-regulated in WT between E10.5 and E14.5 (F)	Genes down-regulated in Tet1-KO PGCs at E12.5 (F)	55	679	14580	3609	1	0.33
Genes up-regulated in WT between E10.5 and E14.5 (F)	Genes up-regulated in Tet1-KO PGCs at E14.5 (F)	26	1070	10770	3609	1	0.07
Genes up-regulated in WT between E10.5 and E14.5 (F)	Genes down-regulated in Tet1-KO PGCs at E14.5 (F)	882	1135	10770	3609	1.98E-231	2.32
Genes down-regulated in WT between E10.5 and E14.5 (F)	Genes up-regulated in Tet1-KO PGCs at E10.5	189	1070	17120	3882	1	0.78
Genes down-regulated in WT between E10.5 and E14.5 (F)	Genes down-regulated in Tet1-KO PGCs at E10.5	836	1695	17120	3882	2.69E-144	2.18
Genes down-regulated in WT between E10.5 and E14.5 (F)	Genes up-regulated in Tet1-KO PGCs at E12.5 (F)	93	192	14580	3882	2.51E-11	1.82
Genes down-regulated in WT between E10.5 and E14.5 (F)	Genes down-regulated in Tet1-KO PGCs at E12.5 (F)	490	679	14580	3882	1.98E-143	2.71
Genes down-regulated in WT between E10.5 and E14.5 (F)	Genes up-regulated in Tet1-KO PGCs at E14.5 (F)	805	1070	10770	3882	1.84E-168	2.09
Genes down-regulated in WT between E10.5 and E14.5 (F)	Genes down-regulated in Tet1-KO PGCs at E14.5 (F)	74	1135	10770	3882	1	0.18

Table 4.2 – Loss of Tet1 promotes increased stability of the demethylated pre-meiotic germ cell state. *Overlap between genes differentially up- or down-regulated in Tet1-KO E10.5, E12.5 male or female, and E14.5 male or female PGCs and genes differentially down-regulated or up-regulated between E10.5 and E14.5 in wild type male or female PGCs. p-value computed by hyper-geometric test.*

The above results suggested that loss of Tet1 leads to initially precocious development of early (E10.5) PGCs followed by delayed progression of late (E14.5) germ cells. To test whether this is indeed the case, I first considered the dynamic changes PGCs undergo between E10.5 and E14.5 in both male and female embryos. First, E10.5 PGCs are subject to a wave of global DNA demethylation, with the bulk of 5mC present at E10.5 lost by E12.5, resulting in a nearly completely demethylated state by E14.5 in both male and female germ cells (see Chapter 3 for details). Immediately following global DNA demethylation, female germ cells enter into meiosis I starting at E13.5 and concurrently down-regulate expression of core pluripotency-associated genes; male germ cells, on the other hand, gradually and asynchronously enter into mitotic arrest between E12.5 and E14.5, and subsequently undergo a wave of extensive *de novo* DNA methylation starting at E15.5.

Considering these observations, DNA demethylation should be initiated earlier in a model whereby Tet1 loss induces the precocious development of E10.5 PGCs, as suggested by the aforementioned transcriptional analysis. To test this prediction, I isolated *Tet1*-KO and wild-type germ cells from embryos between E10.5 and E14.5, and used reduced representation bisulphite sequencing (RRBS) to study locus specific methylation (see Chapter 3 for details) and quantitative liquid chromatography-mass spectrometry (LC/MS) to study global DNA methylation. Consistent with previous reports (Seisenberger et al., 2012), on the predominantly C57B6 genetic background, global DNA demethylation is initiated in wild-type PGCs between E11.5 and E12.5 (Figure 4.5A). In the absence of Tet1, DNA demethylation in PGCs does indeed proceed precociously, with the process being initiated already between E10.5 and E11.5 (Figure 4.5A), a full day earlier than seen in wild-type germ cells. This precocious DNA demethylation is observed in both female and male PGCs. I noticed that the observed hypo-methylation at E10.5 appeared more extensive as determined by RRBS

(representing mainly methylation levels at regulatory elements such as promoters and enhancers) than as determined by LC/MS (representing mainly methylation at repetitive elements) (Figures 4.5B-C). It remains unclear whether this is a sampling artefact (as different sets of embryos were used for RRBS and LC/MS analysis at E10.5); whether demethylation kinetics at regulatory elements are in fact more affected than at repetitive elements; or, most likely, whether this simply reflects the different DNA demethylation kinetics normally observed in wild type PGCs for the uniquely mappable and repetitive parts of the genome (as discussed in Chapter 3). Regardless of this slight discrepancy, both methods (RRBS, LC/MS) confirm precocious DNA demethylation in *Tet1*-KO PGCs at E11.5 and E12.5 (Figure 4.5).

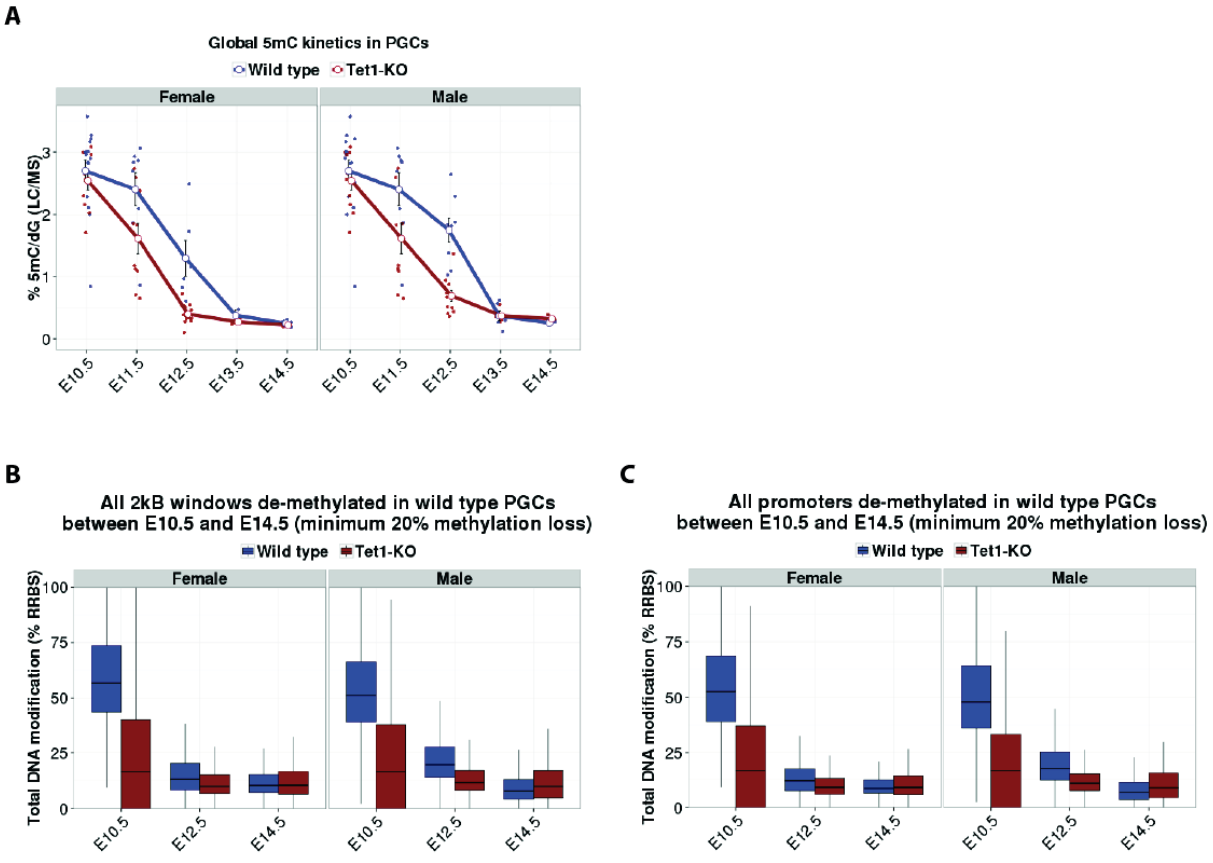


Figure 4.5 – Precocious DNA demethylation is observed in Tet1-KO PGCs. ***A)*** *Dynamics of global DNA demethylation as determined by LC/MS analysis. Each point represents the 5mC levels of PGCs isolated from a single embryo. Mean 5mC levels at a given stage and standard error of the mean are depicted by solid white points with error bars.* ***B)*** *DNA modification levels in E10.5, and E12.5 and E14.5 female (left) and male (right) Tet1-KO and wild type PGCs at 2kB windows genome-wide with a minimum 20% loss of methylation observed in wild type PGCs between E10.5 and E14.5.* ***C)*** *DNA modification levels in E10.5, and E12.5 and E14.5 female (left) and male (right) Tet1-KO and wild type PGCs at all promoters (-1kB/+500bp) with a minimum 20% loss of methylation observed in wild type PGCs between E10.5 and E14.5. Note that for all boxplots: 1) the upper and lower "hinges"*

*correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.*

The observation that precocious development of early (E10.5-E11.5) germ cells is followed by delayed progression of late (E13.5-E15.5) PGCs in the absence of Tet1 predicts one of two distinct phenotypes for late *Tet1*-KO PGCs depending on the sex of the embryos from which they were isolated. In females, loss of Tet1 should lead to delayed transition between mitotic and meiotic cell cycle profiles starting at E12.5-E13.5, concurrent with both delayed induction of genes specifically up-regulated in meiosis I and down-regulation of genes associated with the pluripotency network. Consistently, E14.5 *Tet1*-KO female PGCs clearly show sustained expression of genes associated with the mitotic cell cycle; in wild type germ cells, by comparison, expression of these genes is clearly reduced at this stage (Figure 4.6A). Conversely, meiotic cell cycle genes clearly show weaker activation between E10.5 and E14.5 in *Tet1*-KO PGCs (Figure 4.6A). In addition to showing delayed transition from mitotic to meiotic cell cycling, E14.5 *Tet1*-KO female germ cells additionally have clearly reduced expression of genes specifically up-regulated during meiosis I (Figures 4.6C,E), as well as much higher expression of genes associated with pluripotency (Grskovic et al., 2007), including *Oct4*, *Nanog*, *Prdm14*, and *Sox2* (Figures 4.6B,D). Intriguingly, the expression levels of both pluripotency factors and meiotic genes in *Tet1*-KO E10.5 and E14.5 female PGCs tend to be closer to the levels observed in E12.5 wild type female germ cells than those observed in their stage-matched wild type counterparts (Figures 4.6B-E). Furthermore, it should be noted that, although the expression levels of meiotic genes are reduced relative to wild type levels at E14.5, these genes are nevertheless still clearly up-regulated between E12.5 and E14.5 in *Tet1*-KO female germ cells (Figures 4.6C,E). Thus, expression of meiotic genes can clearly be induced in the absence of Tet1, but appears to occur with delayed kinetics starting at E12.5. Ultimately, Tet1 loss leads to impaired progression through meiosis and extensive apoptotic death at later stages of female foetal germ cell development, as has been previously well characterised (Yamaguchi et al., 2012).

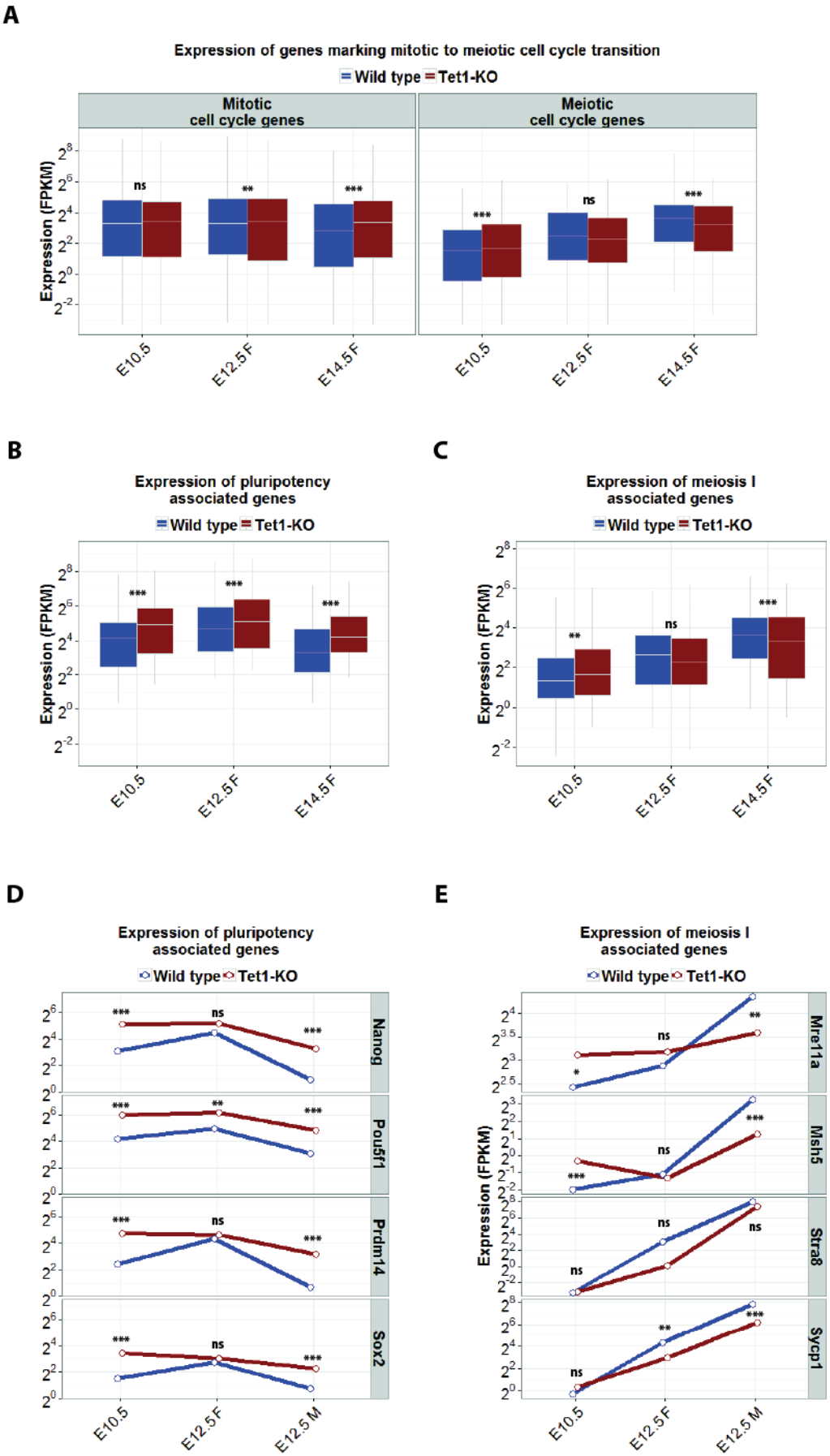

A
Expression of genes marking mitotic to meiotic cell cycle transition
Wild type Tet1-KO
Mitotic cell cycle genes
Meiotic cell cycle genes
Expression (FPKM)
ns
**
***
E10.5
E12.5F
E14.5F
B
Expression of pluripotency associated genes
C
Expression of meiosis I associated genes
D
Nanog
Pou5f1
Prdm14
Sox2
E12.5M
E
Mre11a
Msh5
Stra8
Sycp1

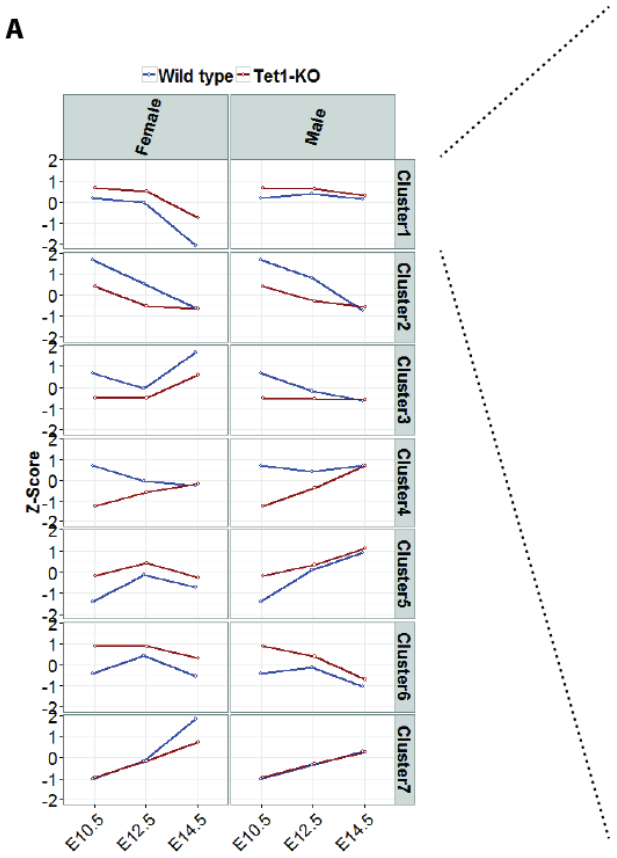
Figure 4.6– Tet1-KO female PGCs show delayed transition between the pre-meiotic and meiotic developmental programmes between E12.5 and E14.5. A) *Expression levels of genes associated with mitotic (left) or meiotic (right) cell cycles in E10.5, E12.5 and E14.5 female Tet1-KO and wild type PGCs. p-values computed by GSEA software. *: p-value < 0.05, **: p-value < 0.01, *** p-value < 0.001.* ***B-C)*** *Expression levels of genes associated with pluripotency (B) and meiosis I (C) in E10.5, E12.5 and E14.5 female Tet1-KO and wild type PGCs. p-values computed by GSEA software. *: p-value < 0.05, **: p-value < 0.01, *** p-value < 0.001.* ***D-E)*** *Expression levels of specific pluripotency associated transcription factors (D) or meiotic genes (E) in E10.5, E12.5 and E14.5 female Tet1-KO and wild type PGCs. adjusted p-values (adj. p-vallue) computed by DESeq software. *: adj. p-value < 0.05, **: adj. p-value < 0.01, *** adj. p-value < 0.001. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.*

In males, delayed progression of late (E13.5-E15.5) PGCs in the absence of Tet1 predicts that fewer *Tet1*-KO germ cells should have entered into mitotic arrest at E14.5-E15.5 and that lower DNA methylation levels should be observed in *Tet1*-KO germ cells at E15.5, immediately following the normal initiation of *de novo* DNA methylation. Considering the difference with respect to the timing of meiosis initiation in females (starting at E13.5) and of mitotic arrest/*de novo* methylation in males (E14.5-E15.5), a much stronger phenotype should be observed in female germ cells compared with male germ cells at E14.5. Indeed, while nearly 20% of genes are differentially expressed in E14.5 female *Tet1*-KO PGCs, less than 5% of genes have significantly altered expression in male *Tet1*-KO germ cells at this time point. Due to time constraints, it remains to be validated whether male germ cells do indeed show delayed entry into mitotic arrest and *de novo* DNA methylation kinetics, and if so, what the exact consequences for male germ cell development are.

Altogether, the observations described above suggest that Tet1 loss results in a phenotype characterised by increased stability of the de-methylated, pre-meiotic germ cell state in PGCs, with the Tet1-KO PGC population seemingly developing precociously at early stages (E10.5-E11.5) of gonadal germ cell development followed by the delayed progression of late (E13.5-E15.5) PGCs.

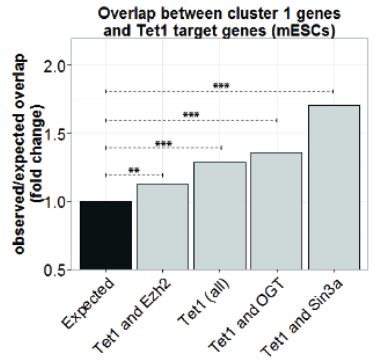
4.2.2 Identification and characterisation of conserved de-regulated Tet1-target genes associated with increased stability of the demethylated pre-meiotic germ cell state

The results presented above suggest that loss of Tet1 in PGCs promotes increased stability of the de-methylated, pre-meiotic germ cell state. Consequently, Tet1 appears to play an important role in regulating the timing of the major transcriptional and developmental transition that occurs in mouse foetal germ line between E10.5 and E14.5, including global DNA demethylation and entry into meiosis. However, the question remains as to how exactly Tet1 mediates progression of the events involved in this transition on the molecular level. To study this, I first had to identify a set of genes likely directly regulated by Tet1 in PGCs. However, as Tet1 loss results in both precocious and delayed transcriptional transitions in PGCs, differential expression analysis at any given stage identifies both direct targets of Tet1 as well as many genes whose altered expression is primarily an artefact of deregulated developmental timing.



D

Cluster 1 enriched GO terms (biological processes)	Fold enrichment	adj. p-value
Mitotic cell cycle	3.7	4.42E-07
Ribosome biogenesis	4.9	3.65E-06
ncRNA processing	4.0	8.99E-06
Response to DNA damage stimulus	2.5	4.38E-03
DNA repair	2.5	3.41E-02



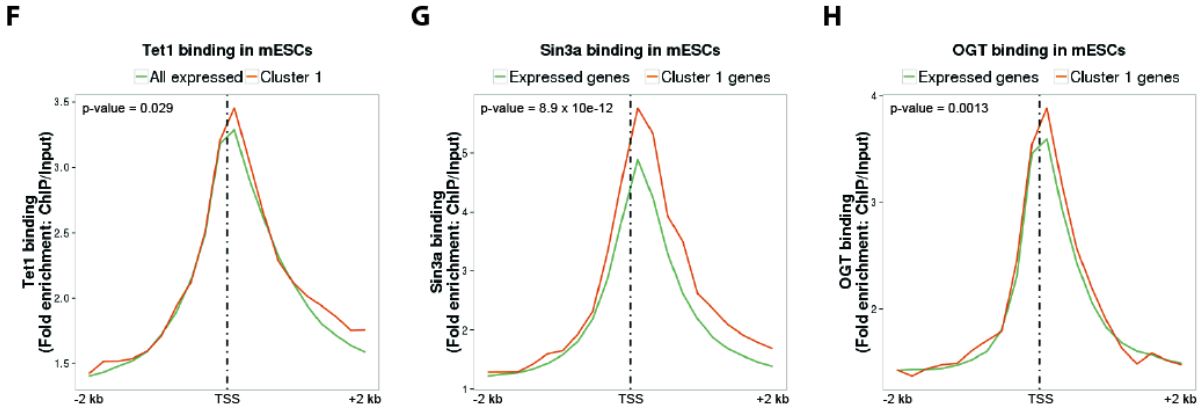


Figure 4.7– Tet1 down-regulates the levels of a conserved cluster of genes associated with the mitotic cell cycle and DNA repair in mouse PGCs and embryonic stem cell (mESCs). A) *K-means clustering identifies 7 clusters of genes differentially expressed in Tet1-KO PGCs in at least one stage of development based on their similar gene expression profiles through wild type PGC development as well as in response to Tet1 loss. The transcriptional profile of each cluster during both female (left) and male (right) PGC development is depicted based on the median Z-score of genes within a given cluster at a given stage. Of note, transcriptional profiles in Tet1-KO (red) and wild type (blue) PGCs are depicted separately.* ***B-C)*** *Gene set enrichment analysis (GSEA) showing enrichment of cluster 1 genes among genes differentially expressed in E13.5 Tet1-KO female PGCs (B) (Yamaguchi et al., 2012) and Tet1-KO mESCs (C) (Hon et al., 2014). Nominal enrichment score (NES) and p-value computed by GSEA software.* ***D)*** *Top gene ontology (GO) biological processes enriched for among cluster 1 genes. Adjusted p-value (adj. p-value) and fold enrichment computed by DAVID.* ***E)*** *Overlap between cluster 1 genes and genes associated with either all Tet1 binding sites or Tet1 binding sites also bound by one of the chromatin regulators Tet1 is known to recruit (Sin3a, OGT, Ezh2). p-value computed by hyper-geometric test.* ***F-H)*** *Mean binding profile for Tet1 (F), Sin3a (G), and OGT (H) in mESCs for genes in cluster 1 (orange) and all expressed genes (green). p-value reflects enrichment of Tet1 (F), Sin3a (G) and OGT (H) at TSS -1kb/+500bp for cluster 1 genes compared to all expressed genes (Mann-Whitney U-test).*

In view of this, I used a two part meta-analysis approach to identify genes likely directly de-regulated by loss of Tet1 in PGCs. First, I identified all genes differentially expressed (adjusted p-value < 0.05, as computed by DESeq2) in *Tet1*-KO PGCs in at least one of the five tested stages (over 5,000 genes in all), and performed K-means clustering to group these genes based on their similar gene expression profiles through wild type PGC development as well as in response to Tet1 loss (Figure 4.7A). Among the most striking features revealed by this clustering analysis was the observation that, while three groups of genes (cluster 1, 5, and 6) all show consistently higher expression in *Tet1*-KO PGCs regardless of stage or sex, no cluster shows the opposite relationship (i.e. consistently lower gene expression across samples), suggesting that Tet1 may predominantly act as a transcriptional repressor in PGCs (Figure 4.7A), which would be consistent with observations in other cell systems (Williams et al., 2011). While this is contrary to the previously proposed primary function for Tet1 in the murine foetal germ line in activating expression of meiotic genes in E13.5 female germ cells (Yamaguchi et al., 2012), it should be noted that comparison to the previously published RNA-Seq data from which these conclusions were drawn confirmed that both cluster 1 (Figure 4.7B, Table 4.3A) and cluster 6 (Table 4.3A) are indeed highly significantly enriched among up-regulated genes in the absence of Tet1 at E13.5.

A Top 10 up-regulated gene sets in Tet1-KO PGCs and mESCs

E10.5 Tet1-KO PGCs - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER5	547	4.397	0.000	0.000
CLUSTER6	460	4.222	0.000	0.000
CLUSTER1	708	2.916	0.000	0.000
TRANSLATION	148	2.171	0.000	0.003
MEIOTIC_RECOMBINATION	15	2.140	0.000	0.004
DNA_REPAIR	115	2.054	0.000	0.008
DNA_RECOMBINATION	39	2.020	0.000	0.009
MEIOSIS_I	18	2.014	0.003	0.008
CLUSTER7	902	1.982	0.000	0.012
RESPONSE_TO_DNA_DAMAGE_STIMULUS	147	1.922	0.000	0.019

E12.5 female Tet1-KO PGCs - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER1	707	3.138	0.000	0.000
CLUSTER6	449	2.896	0.000	0.000
CLUSTER5	547	2.869	0.000	0.000
TRANSLATION	145	2.295	0.000	0.000
RIBOSOME_BIOGENESIS_AND_ASSEMBLY	17	2.016	0.000	0.004
RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS_AND_ASSEMBLY	82	1.986	0.000	0.006
RNA_PROCESSING	163	1.980	0.000	0.006
TRNA_METABOLIC_PROCESS	19	1.971	0.000	0.006
M_PHASE_OF_MITOTIC_CELL_CYCLE	79	1.929	0.000	0.008
REGULATION_OF_MITOSIS	36	1.926	0.000	0.007

E12.5 male Tet1-KO PGCs - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER6	449	2.978	0.000	0.000
TRANSLATION	145	2.542	0.000	0.000
CLUSTER1	707	2.357	0.000	0.000
RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS_AND_ASSEMBLY	82	2.299	0.000	0.000
PROTEIN_RNA_COMPLEX_ASSEMBLY	64	2.175	0.000	0.001
CLUSTER5	547	2.122	0.000	0.001
RNA_PROCESSING	163	2.068	0.000	0.003
RIBOSOME_BIOGENESIS_AND_ASSEMBLY	17	2.018	0.000	0.004
RRNA_METABOLIC_PROCESS	15	1.997	0.003	0.005
RNA_SPLICING	86	1.932	0.000	0.010

E14.5 female Tet1-KO PGCs - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER1	701	5.217	0.000	0.000
CLUSTER6	414	3.720	0.000	0.000
CLUSTER5	513	2.848	0.000	0.000
CELLULAR_LIPID_CATABOLIC_PROCESS	18	2.110	0.000	0.012
LIPID_CATABOLIC_PROCESS	18	2.083	0.000	0.014
AMINO_ACID_CATABOLIC_PROCESS	17	1.898	0.004	0.071
REGULATION_OF_MITOTIC_CELL_CYCLE	18	1.889	0.002	0.063
M_PHASE_OF_MITOTIC_CELL_CYCLE	72	1.889	0.000	0.056
NITROGEN_COMPOUND_CATABOLIC_PROCESS	19	1.876	0.009	0.054
CELLULAR_CATABOLIC_PROCESS	153	1.816	0.000	0.082

E14.5 male Tet1-KO PGCs - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER6	425	1.924	0.000	0.063
TRANSFORMING_GROWTH_FACTOR_BETA_RECEPTOR_SIGNALING_PATHWAY	24	1.793	0.000	0.173
GLYCOPROTEIN_METABOLIC_PROCESS	61	1.766	0.000	0.160
DNA_REPAIR	112	1.714	0.000	0.216
SECONDARY_METABOLIC_PROCESS	17	1.685	0.009	0.240
CLUSTER1	724	1.636	0.000	0.284
ELECTRON_TRANSPORT_GO_0006118	31	1.635	0.005	0.250
CARBOXYLIC_ACID_METABOLIC_PROCESS	115	1.633	0.000	0.227
RESPONSE_TO_DNA_DAMAGE_STIMULUS	137	1.617	0.000	0.236
PROTEIN_IMPORT_INTO_NUCLEUS	35	1.616	0.009	0.217

E13.5 female Tet1-KO PGCs (Yamaguchi et al., 2012) - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER6	419	3.067	0.000	0.000
CLUSTER1	677	2.787	0.000	0.000
REGULATION_OF_CYTOKINE_PRODUCTION	15	1.808	0.004	0.219
GLYCOPROTEIN_METABOLIC_PROCESS	69	1.805	0.000	0.167
CYTOKINE_PRODUCTION	34	1.800	0.000	0.137
APOPTOTIC_NUCLEAR_CHANGES	19	1.714	0.009	0.246
LOCOMOTORY_BEHAVIOR	28	1.714	0.005	0.211
PATTERN_SPECIFICATION_PROCESS	23	1.692	0.011	0.223
NUCLEAR_ORGANIZATION_AND_BIOGENESIS	28	1.690	0.005	0.200
DNA_CATABOLIC_PROCESS	22	1.625	0.030	0.300

Tet1-KO mESCs (Hon et al., 2014) - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
PROTEIN_AMINO_ACID_LIPIDATION	22	1.954	0.000	0.079
LIPOPROTEIN_METABOLIC_PROCESS	28	1.817	0.002	0.219
RNA_SPLICING	88	1.803	0.000	0.169
CLUSTER1	706	1.773	0.000	0.174
MEMBRANE_LIPID_BIOSYNTHETIC_PROCESS	46	1.766	0.002	0.148
RNA_PROCESSING	166	1.745	0.000	0.148
MRNA_PROCESSING_GO_0006397	70	1.718	0.002	0.166
PHOSPHOLIPID_BIOSYNTHETIC_PROCESS	38	1.710	0.004	0.156
PHOSPHOINOSITIDE_BIOSYNTHETIC_PROCESS	23	1.700	0.007	0.151
LIPID_BIOSYNTHETIC_PROCESS	87	1.689	0.000	0.150

B Top 10 down-regulated gene sets in Tet1-KO PGCs and mESCs

E10.5 Tet1-KO PGCs - top negatively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER4	505	-4.199	0.000	0.000
CLUSTER2	1139	-3.847	0.000	0.000
CLUSTER3	630	-3.250	0.000	0.000
PROTEIN_POLYMERIZATION	16	-1.975	0.000	0.014
REGULATION_OF_NUCLEOCYTOPLASMIC_TRANSPORT	18	-1.911	0.000	0.028
REGULATION_OF_INTRACELLULAR_TRANSPORT	20	-1.837	0.000	0.056
SKELETAL_DEVELOPMENT	71	-1.820	0.000	0.061
SYSTEM_DEVELOPMENT	623	-1.750	0.000	0.121
NERVOUS_SYSTEM_DEVELOPMENT	283	-1.741	0.000	0.118
NEURON_DIFFERENTIATION	60	-1.736	0.000	0.111

E12.5 female Tet1-KO PGCs - top negatively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER2	1083	-3.223	0.000	0.000
CLUSTER4	499	-2.225	0.000	0.000
CLUSTER3	623	-2.088	0.000	0.000
SKELETAL_DEVELOPMENT	66	-2.071	0.000	0.001
MUSCLE_DEVELOPMENT	67	-2.023	0.000	0.001
ORGAN_DEVELOPMENT	367	-1.966	0.000	0.001
BRAIN_DEVELOPMENT	37	-1.927	0.000	0.003
SYSTEM_DEVELOPMENT	544	-1.926	0.000	0.003
NEURON_DEVELOPMENT	46	-1.920	0.000	0.003
AXON_GUIDANCE	18	-1.864	0.000	0.007

E12.5 male Tet1-KO PGCs - top negatively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER2	1083	-3.503	0.000	0.000
CLUSTER4	499	-2.597	0.000	0.000
SKELETAL_DEVELOPMENT	66	-2.113	0.000	0.000
MUSCLE_DEVELOPMENT	67	-1.986	0.000	0.004
BRAIN_DEVELOPMENT	37	-1.946	0.000	0.006
CLUSTER3	623	-1.921	0.000	0.008
NEURITE_DEVELOPMENT	38	-1.902	0.000	0.009
HEMOSTASIS	24	-1.889	0.001	0.010
NEURON_DEVELOPMENT	46	-1.889	0.000	0.009
ORGAN_DEVELOPMENT	367	-1.881	0.000	0.009

E14.5 female Tet1-KO PGCs - top negatively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER7	853	-4.924	0.000	0.000
CLUSTER3	561	-4.205	0.000	0.000
MEIOTIC_CELL_CYCLE	27	-2.073	0.000	0.009
REGULATION_OF_DNA_REPLICATION	16	-2.021	0.000	0.013
MEIOSIS_I	17	-1.967	0.002	0.019
DNA_DEPENDENT_DNA_REPLICATION	48	-1.950	0.000	0.018
ER_TO_GOLGI_VESICLE_MEDIATED_TRANSPORT	18	-1.946	0.000	0.016
MEIOTIC_RECOMBINATION	15	-1.863	0.004	0.030
SECRETORY_PATHWAY	60	-1.791	0.000	0.055
SECRETION_BY_CELL	73	-1.733	0.004	0.081

E14.5 male Tet1-KO PGCs - top negatively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
TRANSLATION	134	-2.292	0.000	0.000
RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS_AND_ASSEMBLY	79	-1.902	0.000	0.046
SPLICEOSOME_ASSEMBLY	21	-1.900	0.000	0.031
CELLULAR_BIOSYNTHETIC_PROCESS	224	-1.865	0.000	0.036
PROTEIN_RNA_COMPLEX_ASSEMBLY	61	-1.791	0.000	0.061
MACROMOLECULE_BIOSYNTHETIC_PROCESS	221	-1.781	0.000	0.055
RNA_SPLICINGVIA_TRANSESTERIFICATION_REACTIONS	34	-1.664	0.016	0.130
RNA_PROCESSING	162	-1.573	0.000	0.228
RIBOSOME_BIOGENESIS_AND_ASSEMBLY	17	-1.466	0.054	0.258
MRNA_PROCESSING_GO_0006397	68	-1.460	0.018	0.237

E13.5 female Tet1-KO PGCs (Yamaguchi et al., 2012) - top positively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
CLUSTER7	854	-2.450	0.000	0.000
CLUSTER4	454	-2.078	0.000	0.002
MEIOTIC_CELL_CYCLE	27	-1.776	0.002	0.143
RESPONSE_TO_UV	18	-1.744	0.006	0.145
CLUSTER3	600	-1.705	0.000	0.170
CELLULAR_CARBOHYDRATE_METABOLIC_PROCESS	93	-1.685	0.002	0.170
COENZYME_METABOLIC_PROCESS	33	-1.676	0.007	0.159
POSITIVE_REGULATION_OF_PROTEIN_METABOLIC_PROCESS	38	-1.676	0.005	0.140
DNA_INTEGRITY_CHECKPOINT	19	-1.673	0.009	0.129
DNA_DEPENDENT_DNA_REPLICATION	46	-1.652	0.002	0.137

Tet1-KO mESCs (Hon et al., 2014) - top negatively enriched gene sets	SIZE	NES	NOM p-val	FDR q-val
EXCRETION	20	-1.829	0.004	0.200
MEIOSIS_I	17	-1.762	0.014	0.247
AROMATIC_COMPOUND_METABOLIC_PROCESS	19	-1.750	0.011	0.209
MEIOTIC_CELL_CYCLE	27	-1.685	0.012	0.289
EPIDERMIS_DEVELOPMENT	44	-1.664	0.004	0.284
POSITIVE_REGULATION_OF_TRANSLATION	15	-1.654	0.021	0.265
EXTRACELLULAR_STRUCTURE_ORGANIZATION_AND_BIOGENESIS	16	-1.613	0.029	0.328
CYTOKINE_BIOSYNTHETIC_PROCESS	16	-1.567	0.038	0.370
CYTOKINE_METABOLIC_PROCESS	16	-1.557	0.039	0.363
RESPONSE_TO_BIOTIC_STIMULUS	65	-1.554	0.009	0.340

Table 4.3 – Tet1 down-regulates the expression of a conserved group of genes (cluster 1) associated with the mitotic cell cycle and DNA repair in mouse PGCs and embryonic stem cell (mESCs).

In the second part of the two part meta-analysis, I tested for enrichment of each cluster among genes differentially up- or down-regulated in *Tet1*-KO mouse embryonic stem cells (mESCs) relative to wild type mESCs (Figure 4.7C, Table 4.3A), a normally stable cell system with a transcriptional profile highly similar to PGCs between E10.5 and E13.5 (Grskovic et al., 2007). For comparison, I also tested for enrichment of all GO biological processes as well as for genes associated with known Tet1 binding sites in mESCs (Table 4.3A). The second step of this meta-analysis was performed reasoning that genes directly regulated by Tet1 will likely be enriched among genes de-regulated in both *Tet1*-KO mESCs and PGCs with the same directionality (i.e. up or down-regulated). From this meta-analysis, I was able to identify a group of genes (cluster 1) with consistently higher expression in both PGCs and mESCs lacking Tet1 (Figures 4.7A-C, Table 4.3A). In wild type male germ cells, expression of these genes increases slightly between E10.5 and E12.5 followed by a slight decrease between E12.5 and E14.5 (Figure 4.7A); in wild type female PGCs, on the other hand, these genes show relatively stable median expression between E10.5 and E12.5, followed by a considerable decrease between E12.5 and E14.5 (Figure 4.7A). Importantly, expression of these genes is on average higher in *Tet1*-KO germ cells at all stages tested as well as in mESCs, with the increase in expression most pronounced in E14.5 female germ cells and E10.5 PGCs (Figures 4.7A-C, Table 4.3). Of note, GO analysis revealed that the biological processes most significantly enriched among cluster 1 genes include the mitotic cell cycle and DNA repair (Figure 4.7D), terms whose up-regulation would in theory be consistent with both precocious DNA demethylation (by either an active or passive mechanism) and the delayed development of late PGCs. Remarkably, gene set enrichment analysis revealed that cluster 1 genes are actually more enriched among up-regulated genes in *Tet1*-KO mESCs than all genes associated with known Tet1 binding sites, highlighting the robustness of this meta-analysis for identifying genes whose de-regulated expression is likely a primary consequence of Tet1 loss in PGCs. Tet1 has previously been shown to regulate transcription by directly binding to gene promoters (Huang et al., 2014; Williams et al., 2011), while Tet2 appears to affect gene expression by acting on gene bodies and enhancer elements (Hon et al., 2014; Huang et al., 2014). Indeed, Tet1 levels are particularly elevated around the promoter regions of cluster 1 genes in mESCs (Figure 4.7F), and consistently, a highly significant overlap is observed between cluster 1 genes and

genes associated with Tet1 binding sites, especially those where binding of histone deacetylase complex component Sin3a or O-linked N-acetylglucosamine transferase (Ogt) is also observed (Figures 4.7E-H). These observations further suggest that cluster 1 represents a robust group of Tet1-target genes.

Consistent with the known binding preference of Tet1 for unmethylated CpG-rich regions (Williams et al., 2011), promoters of cluster 1 genes tend to be hypo-methylated in both PGCs (Figure 4.8A) and mESCs (Figure 4.8B); moreover, only very low levels of 5hmC are detected at these regions in both systems (Figures 4.8A-B). Intriguingly, however, cluster 1 genes appear to have a high number of 5-formyl and 5-carboxy modified CpGs around their promoter regions in mESCs, with particularly high levels observed in the 400bp immediately 3' to the transcription start site (TSS) (Figure 4.8B). As 5fC and 5caC are known intermediates of active DNA demethylation (He et al., 2011), I considered the possibility that transcriptional de-regulation of cluster 1 genes may be related to the differential methylation of their promoter regions. However, cluster 1 promoters show no significant overlap with hyper- or hypo-methylated promoters in *Tet1*-KO mESCs (Figure 4.8D) or E10.5 PGCs (Figure 4.8C). In fact, no obvious relationship between differential methylation (either hypo- or hyper-methylation) and differential gene expression could be observed in *Tet1*-KO PGCs at any stage (Table 4.4), including those isolated from E12.5 and E14.5 female embryos, where hyper-methylation of germ line promoters has previously been suggested to be responsible for the lower expression of meiotic genes (Hackett et al., 2013; Yamaguchi et al., 2012). It should also be noted that no detectable levels of 5fC or 5caC are observed in PGCs by either immunostaining or LC/MS (R. Amouroux, personal communication).

Differentially methylated promoter gene sets	Stage used for comparison to Tet1-KO PGC gene expression	SIZE	NES	NOM p-val	FDR q-val
E10_HYPER-METHYLATED_PROMOTER	E10.5	300	-1.135	0.156	0.505
E10_HYPO-METHYLATED_PROMOTER	E10.5	273	-1.380	0.006	0.292
E12_FEMALE_HYPER-METHYLATED_PROMOTER	E12.5 (female)	24	-1.320	0.106	0.210
E12_FEMALE_HYPO-METHYLATED_PROMOTER	E12.5 (female)	51	-0.876	0.692	0.873
E12_MALE_HYPER-METHYLATED_PROMOTER	E12.5 (male)	29	-1.179	0.220	0.382
E12_MALE_HYPO-METHYLATED_PROMOTER	E12.5 (male)	64	-0.670	0.958	0.971
E14_FEMALE_HYPER-METHYLATED_PROMOTER	E14.5 (female)	29	0.991	0.483	0.735
E14_FEMALE_HYPO-METHYLATED_PROMOTER	E14.5 (female)	21	-1.206	0.215	0.585
E14_MALE_HYPER-METHYLATED_PROMOTER	E14.5 (male)	111	-1.116	0.214	0.638
E14_MALE_HYPO-METHYLATED_PROMOTER	E14.5 (male)	104	1.164	0.171	0.457

Table 4.4– Relationship between differential methylation of gene promoters (as determined by RRBS, see Chapter 3 for details) and differential gene expression in Tet1-KO PGCs.

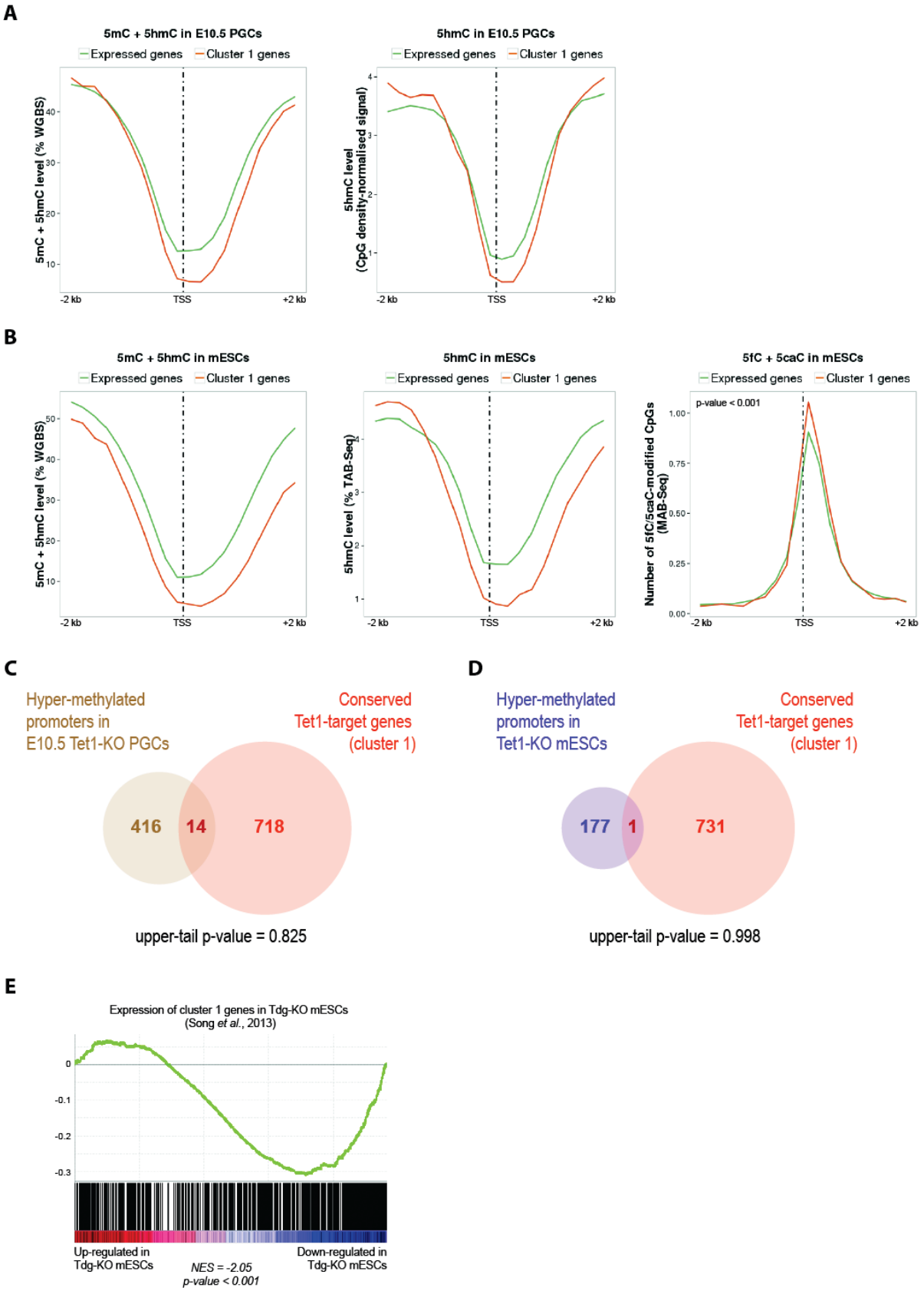


Figure 4.8 – Conserved Tet1 target genes are depleted of 5mC/5hmC but enriched for 5fC/5caC at their transcription start sites. A) *Mean 5mC + 5hmC levels (left) and 5hmC levels (right) around the transcription start sites (TSS) of cluster 1 genes and all expressed genes in E10.5 wild type PGCs.* ***B)*** *Mean 5mC + 5hmC levels (left), 5hmC levels (centre), and number of 5fC/5caC-modified CpGs*

(right) around the transcription start sites (TSS) of cluster 1 genes and all expressed genes in wild type mESCs. p-value refers to the difference in the number of 5fC/5caC-modified CpGs in the 400bp immediately 3' of the TSS in cluster 1 genes compared with all expressed genes (Mann-Whitney U-test). ***C-D)*** *Overlap between cluster 1 genes and genes with differentially hyper-methylated promoters in E10.5 Tet1-KO PGCs (C) or in Tet1-KO mESCs (D). p-values based on hyper-geometric test.* ***E)*** *Gene set enrichment analysis (GSEA) showing enrichment of cluster 1 genes among genes differentially expressed in Tdg-KO mESCs (Song et al., 2013a). Nominal enrichment score (NES) and p-value computed by GSEA software*

In addition to their putative role as intermediates in active DNA demethylation, recent structural analysis has revealed a direct regulatory role for 5fC/5caC sites in slowing down RNA polymerase II (Pol II) transcription rates by promoting Pol II stalling and RNA backtracking specifically at these bases (Kellinger et al., 2012; Wang et al., 2015). In view of the higher levels of 5fC/5caC in the region immediately 3' to the TSS of cluster 1 genes (Figure 4.8B), it is possible that the higher levels of transcription observed for cluster 1 genes is due in part to the higher polymerisation rates of Pol II at these genes in the absence of full 5mC oxidation. Such a model predicts that, in a system with elevated levels of 5fC/5caC, the opposite phenotype should be observed, with cluster 1 genes generally down-regulated. Consistent with such a prediction, expression of cluster 1 genes is highly significantly enriched among down-regulated genes in mESCs depleted of Tdg (Figure 4.8E) (Song et al., 2013a), the glycosylase specifically responsible for the excision of 5fC/5caC from DNA (He et al., 2011). Altogether, these observations suggest that Tet1 potentially regulates transcription of cluster 1 genes via the full oxidation of 5mC to 5fC/5caC at their promoters, and more specifically, immediately 3' to their TSSs. Moreover, considering the very high binding levels of Sin3a (and to a lesser extent, Ogt) observed at cluster 1 promoter regions in wild type mESCs (Figures 4.7E,G,H), reduced recruitment of these two well-characterised Tet1-recruited transcriptional repressors likely further promotes transcriptional up-regulation of these genes in the absence of Tet1.

4.3 Discussion

4.3.1 Re-evaluating the role of Tet1 during transcriptional transitions in mouse foetal germ line

With respect to transcriptional regulation in the mammalian germ line, Tet1 has previously been proposed to primarily act as a transcriptional activator of meiotic genes in late (E12.5-E14.5) female PGCs (Yamaguchi et al., 2012). Here, I show that extensive transcriptional and developmental de-regulation is observed in *Tet1*-KO germ cells well before the initiation of meiosis, in both male and female PGCs. Most strikingly, Tet1 appears to play an important role in regulating the timing of the major developmental transition that occurs in mouse foetal germ line between E10.5 and E14.5, including both global DNA demethylation and entry into meiosis, with Tet1-KO PGCs showing increased stability of the de-methylated, pre-meiotic germ cell state (Figure 4.9).

It has previously been suggested that loss of Tet1 results in failure to demethylate meiotic genes, thus preventing their activation during PGC development (Hackett et al., 2013; Yamaguchi et al., 2012). Although I confirm that the meiotic programme is indeed down-regulated in female germ cells at E14.5 in the absence of Tet1 (Figures 4.6A,C), it should be noted that these genes are not differentially expressed between *Tet1*-KO and wild type female PGCs at E12.5 (Figures 4.6A,C), despite showing considerable activation compared to their levels at E10.5 (Figures 4.6A,C), an observation which is in line with the global wave of DNA demethylation occurring between these two stages (Figure 4.5). It is only in the subsequent transition, between E12.5 and E14.5, that meiotic genes fail to further increase their expression to the same extent as observed in wild type germ cells (Figures 4.6A,C), suggesting that this is likely unrelated to DNA demethylation. Consistently, DNA methylation in E14.5 female *Tet1*-KO PGCs appears largely unaffected at promoter regions, with no

obvious relationship between promoter hyper-methylation and differential gene expression observed (Table 4.3).

Still considering the meiotic phenotype observed in PGCs in the absence of Tet1, it should be noted that E10.5 *Tet1*-KO PGCs show up-regulation of genes normally activated between E10.5 and E14.5 during wild type female germ cell development, including increased expression of genes associated with meiosis I (Figures 4.4, 4.6C). Although the expression levels in E10.5 Tet1-KO PGCs are by no means comparable to those observed in E14.5 wild type female PGCs (Figure 4.6C), these observations do suggest that the initiation of meiotic gene transcription may be precocious in the absence of Tet1. With this in mind, precocious PGC development has previously been described by Peters and colleagues using a conditional knockout of PRC1 components (Yokobayashi et al., 2013). In this study, the authors observed expression of meiotic genes a day earlier than expected, following deletion of PRC1 components at E11.5 (Yokobayashi et al., 2013). Intriguingly, precocious development was associated with an inability to complete meiosis at a later time point (Yokobayashi et al., 2013). Although the authors note that this could be due to a later requirement for PRC1 during female germ cell development, my results are consistent with their model suggesting precocious development ultimately leads to defective meiosis (Yokobayashi et al., 2013). It should be noted, however, that I observed no overlap between genes up-regulated in E12.5 female PRC1-KO PGCs and genes up-regulated in E12.5 female *Tet1*-KO PGCs, suggesting that PRC1 and Tet1 regulate meiotic gene expression via two distinct pathways. Moreover, even if this model is indeed correct, it is unlikely that the entire meiotic defect associated with Tet1 loss is exclusively due to precocious expression of meiotic genes. Rather, this likely represents one part of a more complex regulatory relationship between Tet1 and expression of genes dynamically regulated during entry into meiosis, including notably mitotic cell cycle genes, which are consistently up-regulated in *Tet1*-KO PGCs but which normally require extensive down-regulation upon entry into meiosis (Figure 4.6A, 4.6A-D). It should also be noted that it remains unclear at this point whether Tet1 directly regulates meiotic gene expression, as has been previously proposed (Yamaguchi et al., 2012), or whether reduced expression

of these genes is an artefact of the already de-regulated developmental progression of *Tet1*-KO PGCs. With this is mind, a conditional *Tet1*-KO system may prove invaluable in future studies to decipher the role of Tet1 in meiosis initiation independent of the early confounding transcriptional defects associated with Tet1 loss.

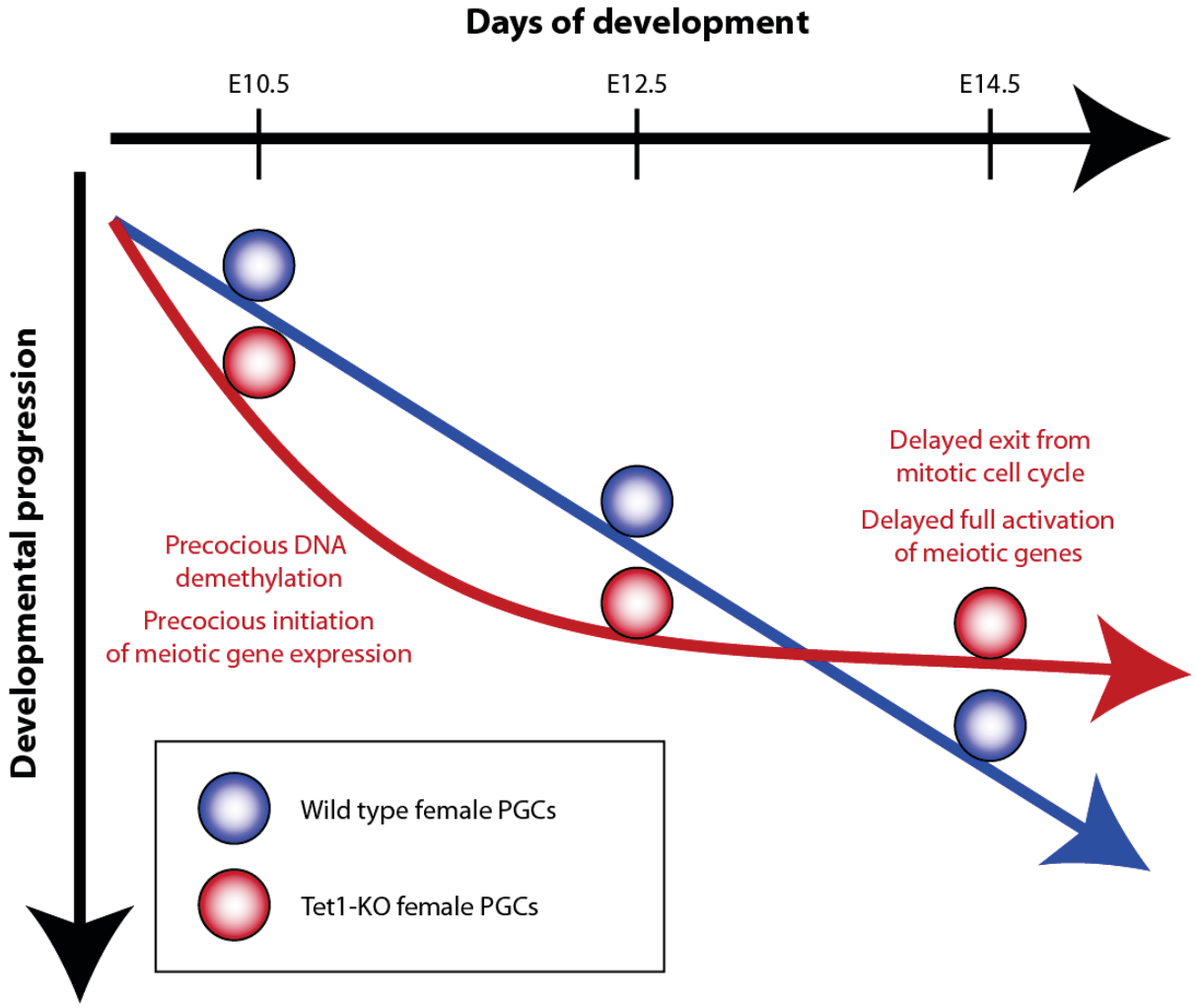


Figure 4.9– Model. *Loss of Tet1 promotes increased stability of the demethylated pre-meiotic germ cell state*

4.3.2 Tet1 regulates transcription via both enzymatic and non-enzymatic activity

Surprisingly, three of the eight identified by clustering meta-analysis comprised genes with consistently higher median expression in *Tet1*-KO PGCs regardless of stage or sex; by comparison, no

cluster showed the opposite relationship, with consistently lower expression of genes (Figure 4.6A). While these results seems in contradiction with the previously proposed model whereby Tet1 acts primarily to activate expression of meiotic genes in female germ cells (Yamaguchi et al., 2012), it is important to reiterate that re-analysis of the RNA-Seq data from which these conclusions were drawn confirmed that two of the three consistently up-regulated clusters I identified are also highly significantly enriched among up-regulated genes in the absence of Tet1 in this previously published dataset (Figure 4.6B). In view these observations, and the now well-characterised role for Tet1 in down-regulating gene expression in mESCs (Kellinger et al., 2012; Vella et al., 2013; Wang et al., 2015; Williams et al., 2011), a role for Tet1 in negatively modulating transcriptional levels appears to represent a potentially overlooked feature of the previously published initial characterisation of Tet1 function in the mammalian germ line.

While I suggest that Tet1 may negatively modulate expression, this should not be interpreted as hypothesizing that Tet1 acts as a transcriptional repressor/silencer; in fact, very few instances of transcriptional activation are observed in *Tet1*-KO PGCs or mESCs for genes normally fully silenced in wild type cells. Rather, Tet1 appears to negatively modulate expression in the sense that it negatively fine-tunes the expression levels of its target genes but does not fully silence them. For example, although the expression of conserved Tet1-target genes (i.e. cluster 1) increases in the absence of Tet1 in both PGCs and mESCs (Figure 4.7A-C, Table 4.2), these genes are already expressed at very high levels on average in wild type cells. As conserved Tet1-target genes are enriched for genes involved in mitotic cell cycling and DNA repair (Figure 4.7D), failure to fine-tune their expression during PGC development appears to lead to increased stability of the demethylated, pre-meiotic germ cell state (Figure 4.9).

At the molecular level, a role for Tet1 in fine-tuning gene expression levels is consistent with the known binding preference of Tet1 for unmethylated CpG-rich regions (Williams et al., 2011), which

show a bias for regulatory elements of predominantly transcriptionally active genes (Weber et al., 2007). Moreover, both the described enzymatic (i.e. 5mC oxidation to 5fC/5caC) and non-enzymatic (i.e. recruitment of Sin3a and Ogt) functions of Tet1 are also consistent with a role in fine-tuning transcription, with 5fC/5caC sites known to reduce RNA polymerase II transcription rates by promoting Pol II stalling and RNA backtracking (Kellinger et al., 2012; Wang et al., 2015), deacetylation of histone proteins inducing a more closed chromatin configuration (Williams et al., 2011; Yang et al., 2002), and glycosylation of transcription factors and the basal transcription machinery known to specifically down-regulate the activity of these components (Vella et al., 2013; Yang et al., 2002). In this respect, the use of a catalytically dead Tet1 mutant will likely prove necessary to decipher the exact contribution of Tet1 enzymatic and non-enzymatic activities for the fine-tuning of gene expression in mouse PGCs in future studies.

Chapter 5

Results

On the role of Akt-mTORC1 signalling during epigenetic reprogramming and licensing for gametogenesis in the mouse foetal germ line

5.1 Introduction

5.1.1 Renewed investigation into the molecular triggers of gonadal DNA demethylation in mouse PGCs

At the onset of this PhD, Tet1 had been proposed as the enzyme initiating gonadal DNA demethylation, including methylation imprint erasure, in mouse PGCs (Hackett et al., 2013; Hajkova et al., 2008; Yamaguchi et al., 2012; Yamaguchi et al., 2013b). However, in view of the results I describe in Chapter 3, DNA demethylation clearly and unambiguously proceeds in the absence of Tet1. As I was unable to detect obvious compensation by the Tet1 homologs Tet2 or Tet3, I suggested in Section 3.3 that investigation into the molecular mechanisms involved in DNA demethylation should be renewed. In this chapter, I take a slightly different approach to understand how the initiation of DNA demethylation is regulated in mouse primordial germ cells, focusing instead on potential extrinsic or intrinsic triggers of this process.

5.1.2 The central role of signalling in mouse PGC development

As described in Chapter 1, signalling is at the heart of murine primordial germ cell development. Specifically, the bone morphogenic protein (BMP) signalling pathway drives specification of PGCs from pluripotent epiblast cells starting at approximately E6.0 (Chang and Matzuk, 2001; de Sousa Lopes et al., 2004; Lawson et al., 1999; Ohinata et al., 2009; Tremblay et al., 2001; Ying et al., 2000; Ying and Zhao, 2001); the Sdf1/Cxcr4 and steel/c-kit pathways mediate PGC migration between E8.5 and E10.5 (Ara et al., 2003; Gu et al., 2009; Molyneaux et al., 2003); and the Fgf9 and (potentially) retinoic acid signalling pathways, among others, appear to initiate the sex specific gonocyte differentiation programmes in male and female germ cells starting at E12.5-E13.5(Anderson et al., 2008; Best et al., 2008; Bowles et al., 2010; Bowles et al., 2006; Colvin et al., 2001; Lin, 2008;

MacLean et al., 2007). The extent of our current understanding regarding the mechanisms by which signalling pathways regulate PGC development is perhaps best reflected by our ability to induce specification (Hayashi et al., 2011; Ohinata et al., 2009), migration (Molyneaux et al., 2003) or sex-specific differentiation (Bowles et al., 2010; Bowles et al., 2006; Lin, 2008) at will in epiblast cells, early PGCs, or post-reprogramming gonadal PGCs, respectively, through the ectopic addition of defined signalling pathway agonists or antagonists. It remains unclear, however, whether epigenetic reprogramming, the other major event observed in early PGC development, is similarly induced in response to PGCs migrating into a new tissue environment, the genital ridge, or whether this process is cell-autonomous, programmed according to some intrinsic clock (McLaren, 2003). In this context, it is interesting to note that reactivation of expression from the inactive X-chromosome appears to be mediated in part by exposure of PGCs to the genital ridge environment, as determined by an elegant flow-cytometry based strategy using transgenic PGCs expressing GFP from the X-chromosome (Chuva de Sousa Lopes et al., 2008).

5.1.3 Current understanding for the role of signalling during gonadal epigenetic reprogramming in mouse PGCs

Recent evidence suggests epigenetic reprogramming may be driven by an extrinsic signal rather than being induced in a cell-autonomous manner (Figure 5.1). As described in Chapter 1, PGC specification can be recapitulated *in vitro* via the two-day differentiation of naïve ES cells into epiblast-like cells (EpiLCs) using FGF2 and activin A, and the subsequent three day induction of EpiLCs into PGC-like cells (PGCLCs) using BMP4, BMP8a, EGF and LIF (Hayashi et al., 2011). However, the PGCLCs generated using this system retain an epigenetic and transcriptional profile similar to pre-gonadal PGCs, including high levels of methylation at imprint control regions (ICRs), as seen prior to imprint erasure, and low expression of methylation sensitive genes such as *Dazl* (Hayashi et al., 2012; Hayashi et al., 2011). Remarkably, however, both ICR demethylation and up-regulation of *Dazl* mRNA expression are observed in PGCLCs upon their re-aggregation with

disaggregated somatic cells of the E12.5 female genital ridge (Hayashi et al., 2012). Moreover, although PGCLCs are reminiscent of early (i.e. E9.5) PGCs at the transcriptional and epigenetic level, they are capable of contributing to spermatogenesis and fertile offspring upon injection into seminiferous tubules of neonatal mice lacking endogenous germ cells (Hayashi et al., 2011). Together, these results suggest the exposure to either the foetal or neonatal gonadal environment may actively trigger erasure of methylation imprints and germ line gene promoters, thus enabling the correct establishment of sex-specific methylation at ICRs and progression through meiosis (Figure 5.1).

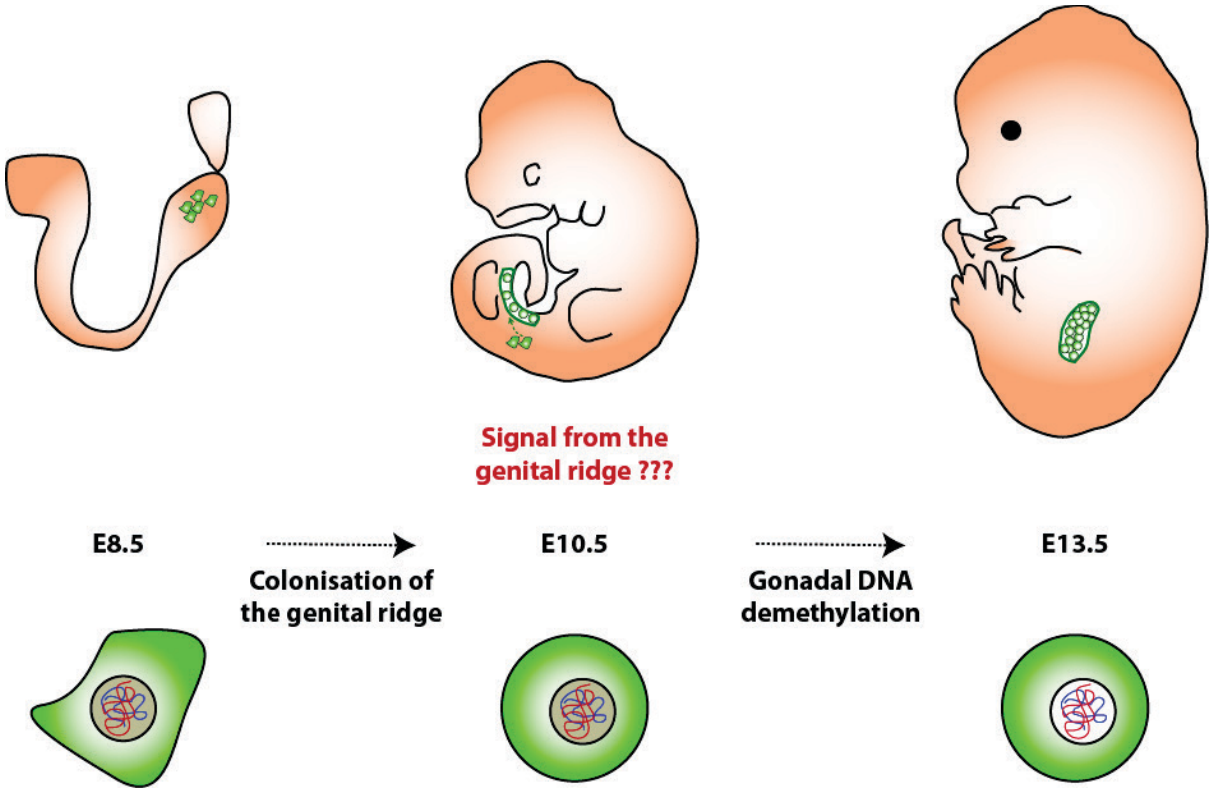


Figure 5.1 – Does the genital ridge environment trigger gonadal DNA demethylation in mouse primordial germ cells?

Additionally, evidence from embryonic germ cell (EGC) derivation further suggests that response to signalling likely regulates gonadal DNA demethylation in mouse PGCs. EGCs are pluripotent stem cells transcriptionally very similar to mouse ESCs (Leitch et al., 2010); however, while ESCs are derived from the inner cell mass (ICM) of the blastocyst, EGCs are derived from PGCs between E8.5

and E12.5 (Matsui et al., 1992; Shovlin et al., 2008; Tada et al., 1998). Traditionally, both ESCs and EGCs have been derived by dissociation of tissue containing PGCs (hindgut for migratory PGCs, genital ridge for post-migratory PGCs) or ICM, followed by culture of the cell suspension in serum-containing medium on feeder cells (Leitch et al., 2013; Shovlin et al., 2008; Tada et al., 1998; Tang et al., 2010). Importantly, despite the fact that both ESCs and EGCs are typically derived from cells with methylation imprints intact, derivation of pluripotent stem cell lines in this way generates ESCs with a methylation status at ICRs similar to that observed in soma, while EGCs typically lack methylation at these regions (Leitch et al., 2013; Shovlin et al., 2008; Tada et al., 1998). It should be noted that, although EGCs derived from E11.5 PGCs completely lack methylation at ICRs (Shovlin et al., 2008; Tada et al., 1998), EGC lines derived from E8.5 PGCs occasionally retain limited methylation at some of these regions (Shovlin et al., 2008); most, however, exhibit a high degree of imprint erasure (Shovlin et al., 2008). More recently, Smith and colleagues developed a method to derive both ESCs and EGCs in completely chemically defined media using the chemical inhibition of two major signalling pathways, MEK signalling and GSK3β signalling (Leitch et al., 2010; Nichols et al., 2009; Ying et al., 2008). Remarkably, EGCs derived from E8.5 PGCs in this way can retain methylation at ICRs to a similar extent as ESCs (Leitch et al., 2013). It is important to note, however, that imprint erasure is still triggered in the majority of 2i-derived EGC cell lines (Leitch et al., 2013). Nonetheless, these observations suggest that loss of methylation imprinting during derivation of EGCs from early PGCs may either be prevented by inhibiting MEK/GSK3β signalling or, alternatively, may actively be triggered by a component of the largely undefined serum/feeder cell media traditionally used.

Considering the above observations implicating response to signalling in erasure of methylation imprints both *in vivo* and *in vitro*, I set out to identify and characterise the major signalling changes occurring within mouse PGCs immediately preceding gonadal epigenetic reprogramming in search of potential triggers of DNA demethylation.

5.2 Results

5.2.1 RNA-Seq profiling between E8.5 and E12.5 reveals transcriptional changes concurrent with the initiation of gonadal DNA demethylation in mouse PGCs

To first address the nature of a potential signal driving gonadal DNA demethylation in mouse primordial germ cells, I carried out RNA-Seq analysis on PGCs isolated from consecutive days of development between E8.5 and E12.5 (see Table 5.1 for details). This was performed on PGCs isolated from embryos with an MF1 x GOF18ΔPE-EGFP (i.e. Oct4-GFP) genetic background, as I had previously established the exact timing of DNA demethylation by high-coverage whole genome bisulphite sequencing and LC/MS (refer to Chapter 3 for details). With respect to the general transcriptional trends observed in wild type foetal germ cells between E8.5 and E12.5, principle component analysis revealed that the major separation between samples occurs between E10.5 and E11.5 (immediately following the major loss of 5mC), with a smaller separation observed in E12.5 female samples, as they begin to initiate the meiosis specific transcriptional programme (Figure 5.2A).

Sample	Total aligned paired end reads	Number of genes detected (FPKM > 0.1)
E8.5 PGCs (replicate 1)	121,905,688	13,537
E8.5 PGCs (replicate 2)	139,588,720	13,537
E9.5 PGCs (replicate 1)	152,193,130	14,197
E9.5 PGCs (replicate 2)	126,826,488	13,564
E10.5 PGCs (replicate 1)	144,600,518	13,216
E10.5 PGCs (replicate 2)	161,042,064	14,901
E11.5 PGCs (replicate 1)	206,981,362	12,936
E11.5 PGCs (replicate 2)	140,954,614	15,276
E12.5 female PGCs (replicate 1)	127,279,268	12,592
E12.5 female PGCs (replicate 2)	138,082,758	17,702
E12.5 male PGCs (replicate 1)	200,924,010	13,712
E12.5 male PGCs (replicate 2)	184,388,882	15,268

Table 5.1 – Summary of RNA-Seq sequencing statistics

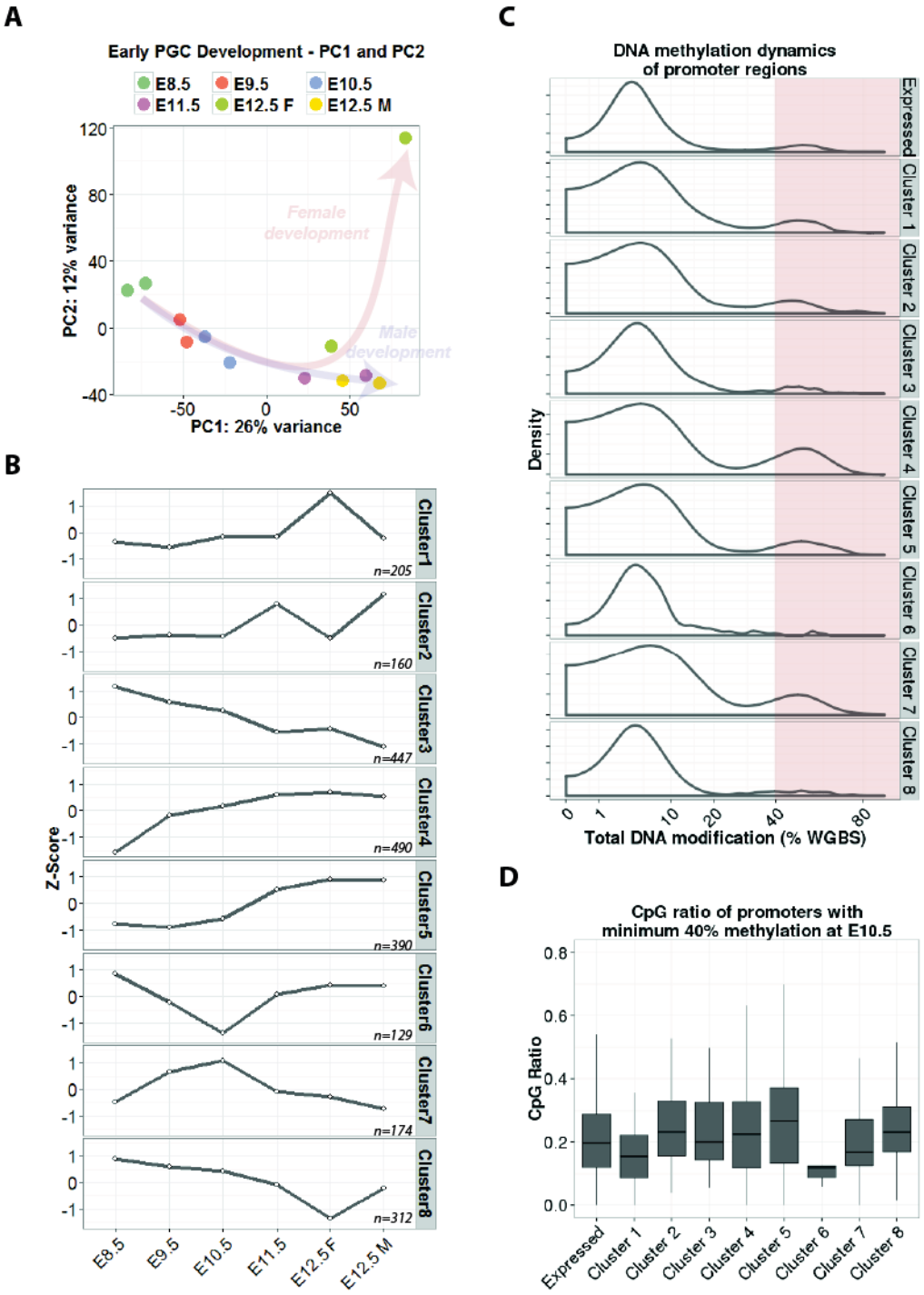


Figure 5.2 – Transcriptional dynamics in mouse PGCs between E8.5 and E12.5. A) *Principle component analysis showing clustering of RNA-Seq samples derived from wild type PGCs between E8.5 and E12.5 (male and female). Blue and pink arrows show normal wild type developmental progression for male and female PGCs, respectively.* ***B)*** *K-means clustering identifies 8 clusters of genes differentially expressed in wild type PGCs between developmental stages. The transcriptional profile of each cluster during PGC development depicted is based on the median Z-score of genes within a given cluster at a given stage. The number of genes in each cluster is indicated.* ***C)*** *Density plot showing distribution of promoter (TSS -1kB/+500bp) methylation in E10.5 PGCs for all genes expressed at E10.5 ("Expressed") and for each of the eight clusters of genes identified in Figure*

5.2B. Methylation levels above 40% are shaded red. ***D)*** *Boxplot showing distribution of promoter (TSS -1kB/+500bp) CpG ratios for all genes expressed in E10.5 PGCs ("Expressed") with a methylated promoter (minimum 40% methylation) and for genes with methylated promoters in E10.5 PGCs (minimum 40% methylation) within each of the eight clusters identified in Figure 5.2B. Note that for all boxplots: 1) the upper and lower "hinges" correspond to the first and third quartiles; 2) the upper whisker extends from the hinge to the highest value that is within 1.5 * IQR (inter-quartile range, or distance between the first and third quartiles); and 3) the lower whisker extends from the hinge to the lowest value within 1.5 * IQR of the hinge.*

To complement this initiation data analysis, I identified all genes differentially expressed between any two stages (over 2,000 genes in all), and performed K-means clustering to identify eight groups of genes based on their similar gene expression profiles through wild type PGC development (Figure 5.2B). Among these, I identified two large groups of genes that show either progressive up-regulation (cluster 4) or down-regulation (cluster 3) of gene expression between E8.5 and E12.5 in both male and female PGCs, as well as a third large cluster that clearly shows transcriptional activation specifically between E10.5 and E11.5, immediately following the major loss of 5mC (cluster 5) (Figure 5.2B). Gene ontology (GO) analysis revealed that the biological processes most enriched for among cluster 3 genes include glycolysis-associated metabolic processes (Table 5.2); those most enriched for among cluster 4 genes include genes involved in the regulation of transcription (Table 5.2); and those most enriched for among cluster 5 genes include those associated with meiosis and mitochondria (Table 5.2). Considering the timing of transcriptional activation (between E10.5 and E11.5) and their enrichment for GO terms associated with meiosis, cluster 5 genes appear to represent the set of genes specifically activated by the wave of gonadal DNA demethylation in PGCs at E10.5 (see Chapter 1 for details) (Karimi et al., 2011; Maatouk et al., 2006). It should be noted, however, that only a limited number of genes in cluster 5 actually have methylated promoters at E10.5, as determined by whole genome bisulphite sequencing (see Chapter 3 for details; Figure 5.2C), and thus transcriptional activation of the bulk of cluster 5 genes is either unrelated to promoter DNA demethylation between E10.5 and E11.5 or is a secondary effect of this event. With respect to the latter possibility, although only a limited number of cluster 5 promoters are methylated at E10.5 (Figure 5.2C), those that are tend to have a higher CpG content than methylated promoters in other

gene clusters (Figure 5.2D). As the significance of promoter methylation for gene repression increases with CpG content (Weber et al., 2007), it may be that transcriptional activation of this small subset of cluster 5 genes with methylated promoters is indeed driven by gonadal DNA demethylation in PGCs, and that their activation subsequently drives transcriptional up-regulation of the remaining cluster 5 genes (thus representing a secondary effect of gonadal DNA demethylation). However, it should be stressed that this remains hypothesis and requires extensive transcriptional network analysis to address in future studies.

Returning to the search for a potential signal driving gonadal DNA demethylation in mouse primordial germ cells, clustering analysis identified two groups of genes potentially implicated in the initiation of this process: cluster 7, which represents genes transiently up-regulated at E10.5; and cluster 6, which represents genes transiently down-regulated at E10.5. However, these clusters are small (174 and 129 genes, respectively), and not statistically significantly enriched for any GO term (Table 5.2). Thus, to date, transcriptional analysis has revealed no obvious candidate signals potentially triggering gonadal DNA demethylation in PGCs, although it should be noted that this is still an ongoing process, with other resources than Gene Ontology currently being investigated.

Cluster 1 - enriched GO terms	p-value	Fold enrichment	adj. p-value
Female gamete generation	1.23E-03	10.5	0.71
Oogenesis	4.22E-03	12.1	0.88
Ferric iron binding	6.07E-03	25.1	0.78
Gamete generation	6.71E-03	3.2	0.89
Reproductive process in a multicellular organism	7.46E-03	2.9	0.85
Cluster 2 - enriched GO terms	**p-value**	**Fold enrichment**	**adj. p-value**
Regulation of cell proliferation	1.35E-03	3.0	0.72
Embryonic morphogenesis	8.70E-03	3.1	0.98
Microtubule binding	9.72E-03	9.0	0.93
NADP or NADPH binding	9.92E-03	19.6	0.74
Tubulin binding	1.76E-02	7.2	0.80
Cluster 3 - enriched GO terms	**p-value**	**Fold enrichment**	**adj. p-value**
Monosaccharide metabolic process	1.10E-09	5.6	1.72E-06
Hexose catabolic process	4.06E-09	11.7	3.17E-06
Glucose catabolic process	4.06E-09	11.7	3.17E-06
Carbohydrate catabolic process	5.55E-09	8.8	2.89E-06
Hexose metabolic process	5.86E-09	5.7	2.28E-06
Cluster 4 - enriched GO terms	**p-value**	**Fold enrichment**	**adj. p-value**
Regulation of transcription	1.59E-07	1.8	2.10E-04
Regulation of transcription, DNA-dependent	8.19E-07	1.9	5.39E-04
Regulation of RNA metabolic process	1.35E-06	1.9	5.94E-04
Zinc ion binding	1.30E-04	1.6	5.16E-02
Transcription regulator activity	3.00E-04	1.8	5.93E-02
Cluster 5 - enriched GO terms	**p-value**	**Fold enrichment**	**adj. p-value**
Mitochondrion	2.49E-06	2.1	6.78E-04
M phase of meiotic cell cycle	3.98E-06	6.9	5.06E-03
Meiosis	3.98E-06	6.9	5.06E-03
Meiotic cell cycle	4.88E-06	6.8	3.11E-03
DNA binding	4.36E-05	1.7	1.68E-02
Cluster 6 - enriched GO terms	**p-value**	**Fold enrichment**	**adj. p-value**
Positive regulation of fibroblast proliferation	2.40E-03	39.8	0.81
Regulation of fibroblast proliferation	7.96E-03	21.9	0.93
Apical part of cell	1.12E-02	5.7	0.81
FAD binding	1.23E-02	8.2	0.93
Organelle outer membrane	1.82E-02	7.1	0.74
Cluster 7 - enriched GO terms	**p-value**	**Fold enrichment**	**adj. p-value**
Cytoskeletal protein binding	5.23E-04	3.8	0.10
Cytoskeleton	4.04E-03	2.2	0.49
Microtubule	4.06E-03	4.6	0.29
Peroxisomal membrane	1.16E-02	18.0	0.48
Microbody membrane	1.16E-02	18.0	0.48
Cluster 8 - enriched GO terms	**p-value**	**Fold enrichment**	**adj. p-value**
Translation factor activity, nucleic acid binding	1.55E-03	5.6	0.42
Glial cell development	3.54E-03	12.8	0.99
Gliogenesis	5.31E-03	7.1	0.96
ncRNA processing	5.72E-03	3.7	0.90
ncRNA metabolic process	6.27E-03	3.3	0.85
Translation initiation factor activity	7.94E-03	6.3	0.75

Table 5.2 – Top GO terms enriched for among gene expression clusters in PGCs (see Figure 5.2B for details).

5.2.2 A signalling screen identifies activation of the Akt-mTORC1 signalling axis in PGCs during gonadal epigenetic reprogramming and licensing for gametogenesis

As transcriptional analysis revealed no obvious candidate signals, I next considered alternative approaches to identify potential signals instigating global DNA demethylation in gonadal PGCs. With this in mind, I used a targeted immunostaining approach to study changes in the activation status of important signalling nodes in PGCs between E8.5 and E11.5 (Figure 5.3). Unexpectedly, this analysis revealed that PGCs are negative for BMP signalling (Figure 5.3B; n=0/30, as determined by phosphorylated Smad1/5) and MEK signalling (Figure 5.3C; n=1/328, as determined by phosphorylated Erk1/2) between these stages. In contrast, PGCs are positive for Wnt signalling (Figure 5.3D; as determined by nuclear localisation of de-phosphorylated β-catenin) at E8.5; however, by E10.5, active β-catenin re-localises to the cytoplasm, at which point it shows intense staining for cytoplasmic/plasma membrane foci (Figure 5.3D). Finally, PGCs show activation of Akt (Figures 5.3A, 5.5A; as determined by phosphorylation of Akt1) between E8.5 (n=1/58) and E10.5 (n=103/205), upon entry into the genital ridge.

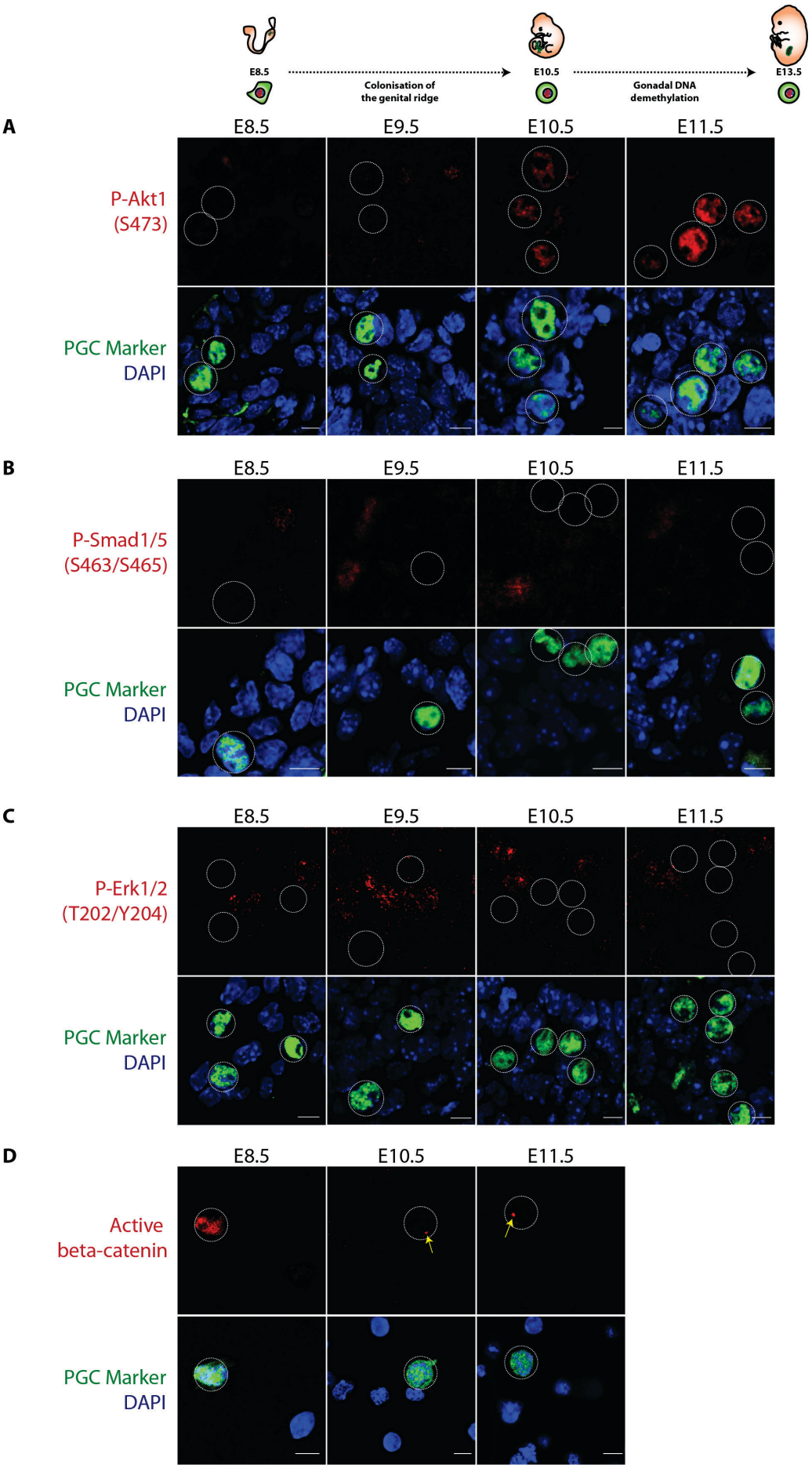

E8.5
Colonisation of the genital ridge
E10.5
Gonadal DNA demethylation
E13.5
A
E8.5
E9.5
E10.5
E11.5
P-Akt1 (S473)
PGC Marker DAPI
B
E8.5
E9.5
E10.5
E11.5
P-Smad1/5 (S463/S465)
PGC Marker DAPI
C
E8.5
E9.5
E10.5
E11.5
P-Erk1/2 (T202/Y204)
PGC Marker DAPI
D
E8.5
E10.5
E11.5
Active beta-catenin
PGC Marker DAPI

Figure 5.3 – Activation status of major signalling pathways during epigenetic reprogramming in mouse PGCs. A-D) *Immunofluorescence analysis for the activation status of Akt signalling (A, as determined by phosphorylation of Akt1), BMP signalling (B, as determined by phosphorylation of Smad1/5), MEK signalling (C, as determined by phosphorylation of Erk1/2) and Wnt signalling (D, as determined by localisation of de-phosphorylated (active) beta-catenin) in PGCs found in the hindgut (E8.5-E9.5), embryonic trunk (E10.5) and genital ridge (E11.5). PGCs are marked by Oct4 (A-C) or GFP (D). Scale bar represents 10 µm.*

The observations described above are of interest for a number of reasons. First, the lack of dynamic changes in either BMP or Erk1/2 signalling likely excludes the possibility that changes in the activation status of these pathways triggers gonadal DNA demethylation; however, it does not exclude the possibility that the absence of their activation in PGCs is a necessary condition for this process to occur. Second, the absence of BMP signalling is particularly unexpected considering that BMP4 signalling via Smad1/5 phosphorylation is absolutely necessary for PGC induction from epiblast cells at the time of PGC specification (E6.25) (Chang and Matzuk, 2001; de Sousa Lopes et al., 2004; Lawson et al., 1999; Tremblay et al., 2001; Ying et al., 2000; Ying and Zhao, 2001). In this context, it should be noted that *Id1*, *Id2*, and *Id3*, which are direct targets of BMP signalling (Miyazono and Miyazawa, 2002), are expressed at high levels in PGCs. This suggests that either their expression has been activated by another mechanism or that the antibody I used is not sensitive enough to detect P-Smad1/5 in PGCs (although clear signalling can be detected in surrounding somatic cells). Further investigation is required to interpret these initial results.

Third, the lack of Erk1/2 signalling is in direct contrast to a previous report which suggested that steel/kit signalling mediates PGC cell survival specifically via the MEK pathway (De Miguel et al., 2002). However, as this previous work is based on the *in vitro* culture of PGCs on feeder cells (De Miguel et al., 2002) while the work I present is based on immunostaining of a significant number (n=1/328) of PGCs in an *in vivo* context, it may be that a requirement for Erk1/2 activation is an artefact of *in vitro* cell culture and does not reflect the *in vivo* reality. Indeed, my observations are consistent with an independent recently published study reporting inhibition of Erk1/2 signalling in

pre-gonadal PGCs *in vivo* (Grabole et al., 2013). Moreover, the absence of Erk1/2 signalling *in vivo* is particularly interesting considering that clear nuclear localisation of active beta-catenin is also observed in E8.5 PGCs. Pluripotent mouse embryonic stem cells (mESCs) and germ cells (mEGCs) can be maintained in the naïve state of pluripotency by dual inhibition of the MEK and GSK3β kinases (Leitch et al., 2010; Nichols et al., 2009; Ying et al., 2008), which respectively result in inhibition of Erk1/2 signalling and nuclear localisation of active beta-catenin. It is interesting to consider the possibility that a similar relationship between signalling pathway activation status and pluripotency transcription factor expression may be observed *in vivo* in PGCs following specification.

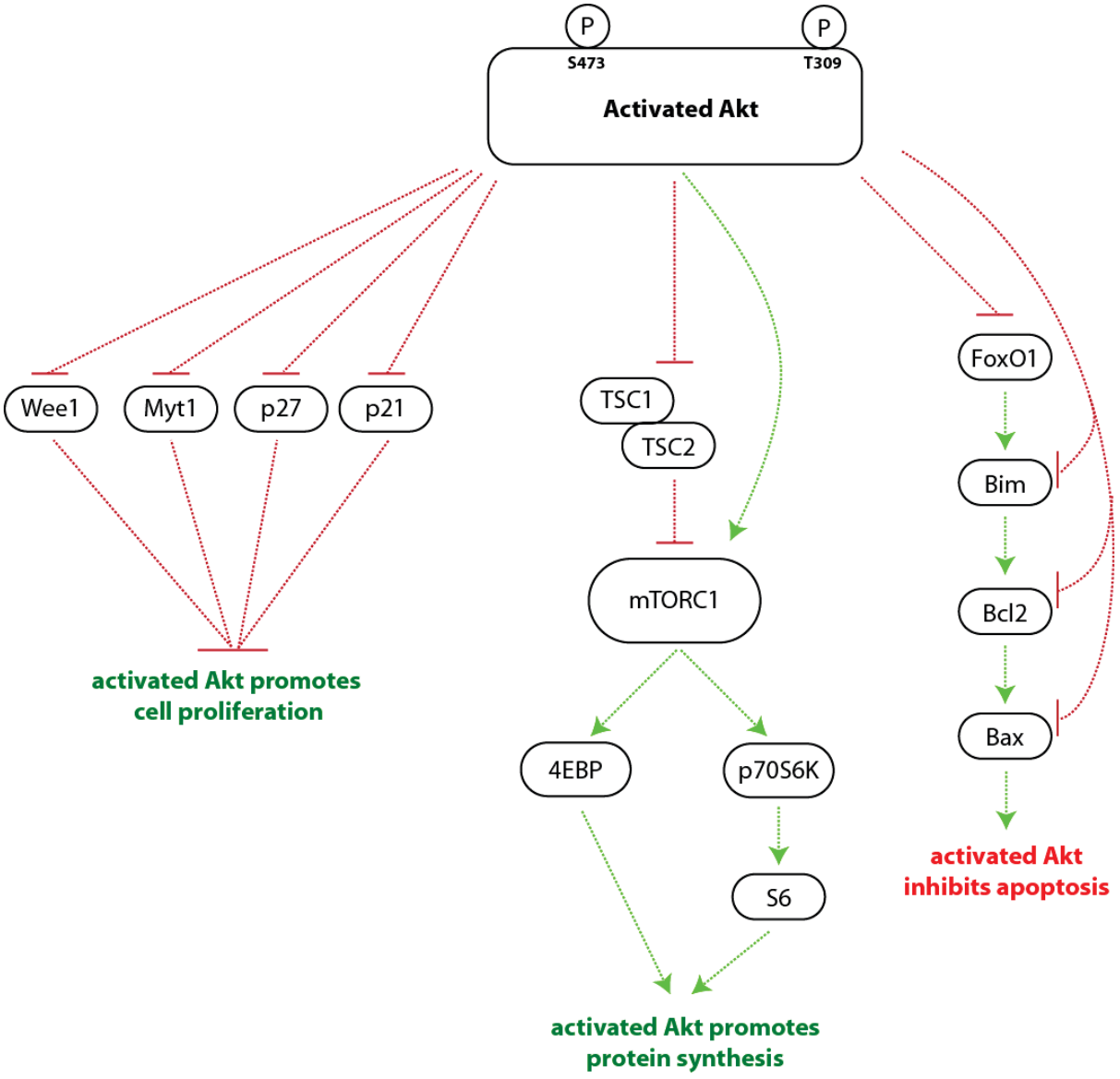


Figure 5.4 – Overview of selected downstream pathways of active Akt signalling.

Most strikingly, the transient activation of Akt signalling in PGCs (E10.5) coincides precisely with the onset of gonadal DNA demethylation. Akt is a major signalling hub, integrating many upstream signals to a number of downstream effectors: induction of proliferation, protection from apoptosis, and stimulation of translation are among the best characterized of these downstream pathways (Figure 5.4) (Manning and Cantley, 2007). In PGCs, the increase in the percentage of cells with Akt signalling does not correlate with an increase in the percentage of cells showing staining for phosphorylated histone 3 serine 10, a marker for mitosis (Figure 5.5D); these results thus suggest that Akt signalling is not driving proliferation of PGCs following their entry into the genital ridge. That said, it should be noted that a significant increase in PGC numbers is observed upon entry into the genital ridge, suggesting that proliferation rates may have actually increased in PGCs between E9.5 and E11.5, in contrast to what is suggested by H3S10P staining. In this context, it is possible that the time in which PGCs spend in different parts of the cell cycle changes through development, and thus H3S10P stainings do not accurately reflect the differences in the rate of PGC proliferation between stages. Further work is needed to interpret these initial results.

In contrast, Akt signalling correlates with the nuclear exclusion of Foxo1 in PGCs (Figure 5.5C). Akt inhibits apoptosis via a variety of mechanisms; among the best characterized is via the direct inhibitory phosphorylation of the apoptotic protein Bax and phosphorylation of the Foxo family of apoptotic transcription factors (Foxo1, Foxo3, and Foxo4), leading to their nuclear exclusion (Tzivion et al., 2011; Yamaguchi and Wang, 2001). Foxo1 shows strong nuclear localisation (44%, n=34) in PGCs upon entry into the genital ridge and concurrent with the initiation of gonadal DNA demethylation (E10.5), but much weaker nuclear localisation in the days immediately preceding (E9.5; 19%, n=31) and (especially) following (E11.5; 5%, n=60) (Figure 5.5C). A possible explanation for these observations is that PGCs are exposed to a source of stress (such as DNA damage, which may be linked to the activation of the base excision repair pathway during active DNA demethylation (Hajkova et al., 2010)), leading to nuclear accumulation of Foxo1. In this context, Akt signalling may be needed for the nuclear exclusion of Foxo1 and thus PGC survival. This hypothesis

has yet to be confirmed and requires further investigation. Moreover, it will be interesting to test whether Foxo3 and Foxo4 show a similar pattern of localisation, and whether phosphorylation of Bax proteins coincides with Akt1 phosphorylation.

Finally, Akt signalling correlates strongly with the activation of cap-dependent translation (Figure 5.5B; as determined by phosphorylation of 4E-BP). During cap-dependent translation, the mRNA 5'-cap is recognized by the translation initiation factor eIF4E (Manning and Cantley, 2007). Normally, the hypo-phosphorylated form of eIF4E binding protein (4E-BP) binds to eIF4E, prevents its interaction with eIF4G, and thus impairs cap-dependent translation initiation; conversely, when hyper-phosphorylated, 4E-BP releases eIF4E, allowing activation of cap-dependent translation. Critically, phosphorylation of 4E-BP is stimulated by the Akt-mTORC1 signalling axis (Figure 5.4) and represents a rate limiting step in cap-dependent translation under most circumstances (Manning and Cantley, 2007). In PGCs, phosphorylation of 4E-BP is distinctly absent in E9.5 PGCs (Figure 5.5B, n=7/51). However, at E10.5, upon entry of PGCs into the genital ridge, strong nuclear and cytoplasmic staining for phospho-4EBP is observed (Figure 5.5B, n=42/58). Thus, the dynamics of 4E-BP phosphorylation correlate well with the dynamics of both Akt signalling and of gonadal DNA demethylation.

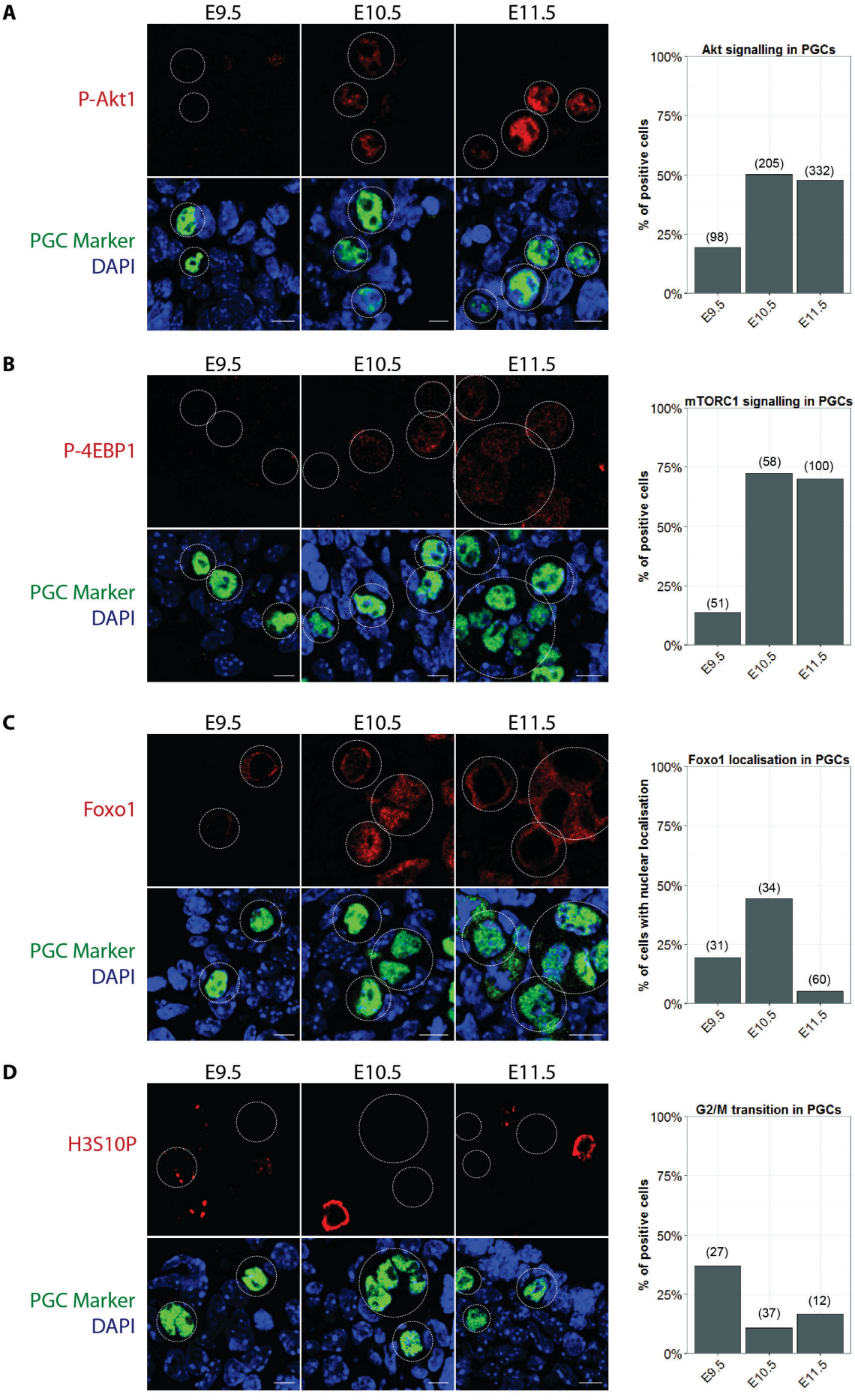

A
E9.5
E10.5
E11.5
P-Akt1
PGC Marker
DAPI
Akt signalling in PGCs
% of positive cells
100%
75%
50%
25%
0%
(98)
(205)
(332)
E9.5
E10.5
E11.5
B
E9.5
E10.5
E11.5
P-4EBP1
PGC Marker
DAPI
mTORC1 signalling in PGCs
% of positive cells
100%
75%
50%
25%
0%
(51)
(58)
(100)
E9.5
E10.5
E11.5
C
E9.5
E10.5
E11.5
Foxo1
PGC Marker
DAPI
Foxo1 localisation in PGCs
% of cells with nuclear localisation
100%
75%
50%
25%
0%
(31)
(34)
(60)
E9.5
E10.5
E11.5
D
E9.5
E10.5
E11.5
H3S10P
PGC Marker
DAPI
G2/M transition in PGCs
% of positive cells
100%
75%
50%
25%
0%
(27)
(37)
(12)
E9.5
E10.5
E11.5

Figure 5.5 – Activation status of downstream targets of Akt pathway identifies Akt-mTORC1 axis as a potential regulator of early gonadal PGC development in mouse. ***A)*** *Immunofluorescence analysis for activation status of Akt signalling (as determined by phosphorylation of Akt1) in PGCs between E9.5 and E11.5. Percentage of cells positive for active Akt signalling at each stage depicted (right), with total number of cells imaged in a given stage indicated above each bar.* ***B)*** *Immunofluorescence analysis for activation status of mTORC1 signalling (as determined by phosphorylation of 4EBP1) in PGCs between E9.5 and E11.5. Percentage of cells positive for active mTORC1 signalling at each stage depicted (right), with total number of cells imaged in a given stage indicated above each bar.* ***C)*** *Immunofluorescence analysis to determine Foxo1 localisation in PGCs between E9.5 and E11.5. Percentage of cells with nuclear Foxo1 staining at each stage depicted (right), with total number of cells imaged in a given stage indicated above each bar.* ***D)*** *Immunofluorescence analysis to determine H3S10 phosphorylation (a marker for cells in G2/M phase) in PGCs between E9.5 and E11.5. Percentage of cells staining positive for H3S10P at each stage depicted (right), with total number of cells imaged in a given stage indicated above each bar. PGCs in the hindgut (E9.5), embryonic trunk (E10.5) and genital ridge (E11.5) are marked by Oct4 for figures A-D. Scale bar represents 10 μm.*

5.2.3 Inhibition of mTORC1 signalling does not impair global DNA demethylation but causes significantly reduced expression of Dazl protein

The timing of Akt-mTORC1 activation hinted at a potential functional role for this signalling pathway in regulating the epigenetic reprogramming and gene expression transition events that are observed in PGCs as they colonise the gonads. To test this hypothesis, I dissected out urogenital ridges from E10.5 embryos and cultured them for 24 or 48 hours in the presence of ethanol (control) or rapamycin (a highly specific and sensitive inhibitor of the mTORC1 complex). Of note, I used C57B6 x GOF18ΔPE-EGFP (i.e. Oct4-GFP) embryos for these experiments as gonadal DNA demethylation appears to initiate slightly later on this genetic background (between E11.5 and E12.5; see Chapter 4 for details), providing a longer incubation time for rapamycin treatment to take effect. Importantly, immunostaining confirmed significant reduction in mTORC1 signalling upon treatment with rapamycin ($p < 0.001$, binomial test), with very few PGCs positive for the phosphorylated form of 4EBP (3%, n=222) compared with ethanol control (39%, n=295).

To first address the importance of mTORC1 signalling for gene expression changes during PGC reprogramming, urogenital ridges were fixed for immunostaining after 24 hrs culture (1 μM rapamycin) with an antibody against the key meiotic regulator Dazl, whose expression specifically increases in PGCs following gonadal DNA demethylation. Consistent with a functional role for Akt-mTORC1 signalling in the regulation of Dazl expression, I observed a highly significant reduction in the number of PGCs expressing Dazl following rapamycin treatment compared with control ($p < 0.001$, binomial test; control: 35% (n=291); rapamycin: 20% (n=314)). Surprisingly, however, PGCs isolated from urogenital ridges cultured under both conditions showed similar global DNA demethylation dynamics, as determined by LC/MS. Specifically, and consistent with global DNA demethylation initiating between E11.5 and E12.5 on this genetic background, limited loss of global 5mC was observed in PGCs cultured in both conditions after 24 hrs (1 μM rapamycin); however, after 48hrs culture (200 nM rapamycin), global DNA demethylation had clearly initiated in both conditions, although considerable 5mC levels are still observed at this time point suggesting that this is an ongoing process.

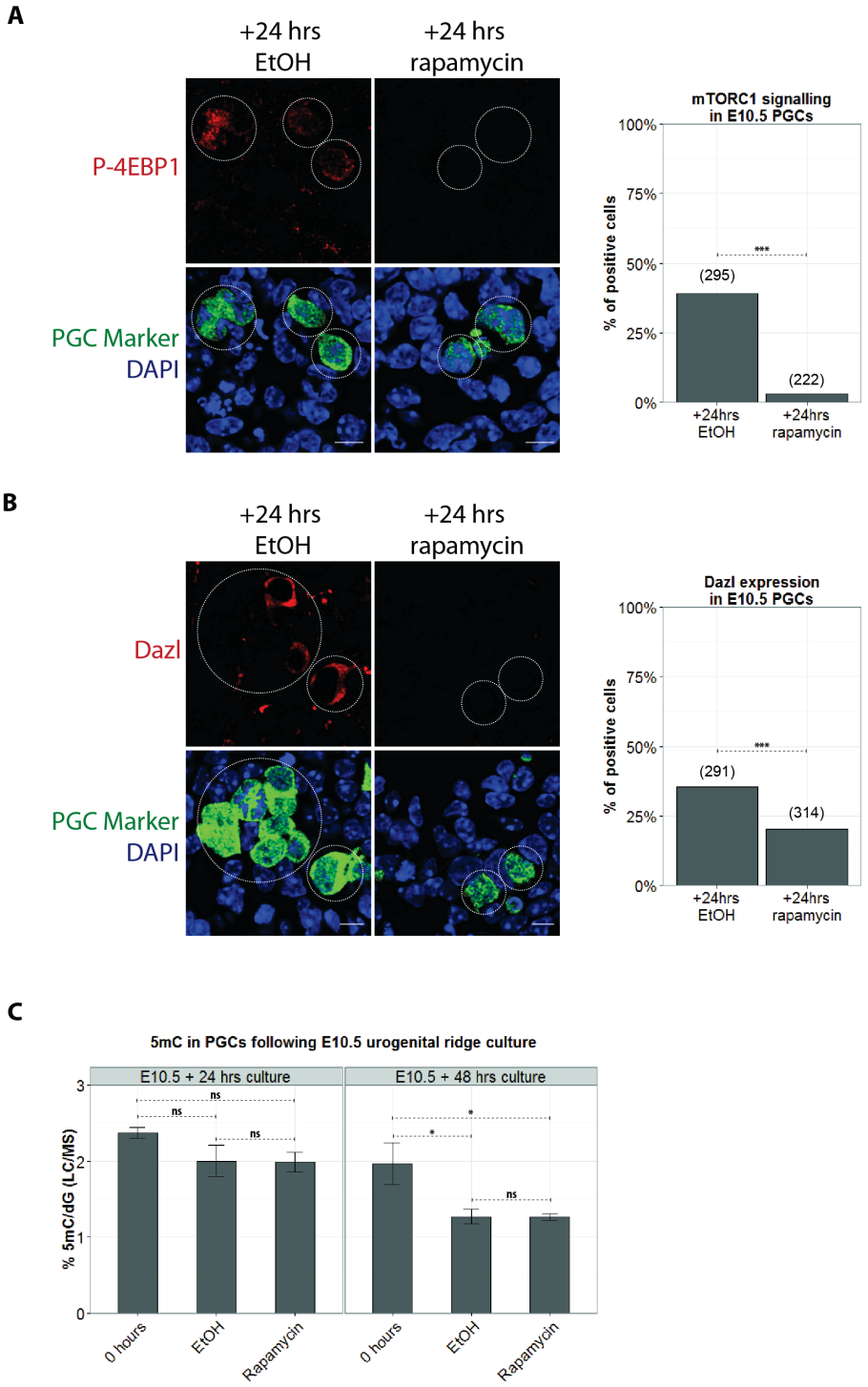


Figure 5.6 - Inhibition of mTORC1 signalling at E10.5 does not impair global DNA demethylation but results in significantly fewer PGCs expressing Dazl protein after 24 hours culture. A) *Immunofluorescence analysis for activation status of mTORC1 signalling (as determined by phosphorylation of 4EBP1) in E10.5 PGCs cultured for 24 hours with either 1 μM rapamycin or with ethanol (control). Percentage of cells positive for active mTORC1 signalling in each condition depicted (right), with total number of cells imaged in a given stage indicated above each bar. p-values are based on binomial test; *: p-value < 0.05, **: p-value < 0.01, *** p-value < 0.001.* ***B)*** *Immunofluorescence analysis for expression of Dazl protein in E10.5 PGCs cultured for 24 hours with either 1uM rapamycin or with ethanol (control). Percentage of cells positive for Dazl expression*

*in each condition depicted (right), with total number of cells imaged in a given stage indicated above each bar. p-values are based on binomial test; *: p-value < 0.05, **: p-value < 0.01, *** p-value < 0.001. PGCs in the genital ridge are marked by GFP for figures A-B.* ***C)*** *Total 5mC levels in E10.5 PGCs cultured for 24 hrs (with either 1 μM rapamycin or with ethanol (control)) or 48 hrs (with either 200 nM rapamycin or with ethanol (control)). p-values based on Student's t-test; ***: p-value < 0.001, **: p-value < 0.01, *: p-value < 0.05. Scale bar represents 10 μm. Bars represent mean value and error bars represent standard error of the mean.*

5.3 Discussion

In this chapter, I present my initial investigation into potential Tet1-independent triggers of gonadal DNA demethylation in mouse PGCs; with this in mind, it should be stressed that this work remains preliminary and is still ongoing. Nonetheless, I reveal: 1) that Akt-mTORC1 signalling is clearly activated concurrent with the initiation of gonadal DNA demethylation; and 2) that inhibition of this pathway using a highly selective and sensitive inhibitor of the mTORC1 complex clearly results in reduced activation of the key meiotic regulator Dazl. To my knowledge, this represents the first description of Akt signalling activation in PGCs in an *in vivo* context, although it should be mentioned that activation of Akt signalling in PGCs has previously been described during both the *in vitro* culture of PGCs on feeder cells and during EG derivation (De Miguel et al., 2002; Moe-Behrens et al., 2003). While my initial results from organ culture experiments suggest that gonadal DNA demethylation is not triggered by activation of Akt-mTORC1 signalling, it should be noted that the LC/MS analysis used to assess DNA methylation mainly reflects 5mC levels at repetitive elements, and thus it remains possible that activation of this pathway may be required specifically for initiating DNA demethylation within the uniquely mappable part of the genome. Future studies using urogenital ridge culture experiments with higher rapamycin concentration and longer incubation time (i.e. 72 hrs, 1 μM rapamycin) coupled with base pair resolution analysis of DNA modification levels will be necessary to fully understand the role of Akt-mTORC1 during gonadal DNA demethylation in PGCs. Such experiments would have the additional benefit of providing information regarding the methylation state of the *Dazl* promoter in late stage PGCs following rapamycin treatment. Specifically, it is possible that the reduced activation of Dazl expression may be driven by reduced

transcription of the *Dazl* locus as a result of impaired DNA demethylation of its promoter region; alternatively, *Dazl* promoter DNA demethylation and transcription may be unaffected by lack of Akt-mTORC1 signalling, but rather a failure to efficiently translate *Dazl* mRNA upon rapamycin treatment may cause the observed reduction in the number of PGCs expressing Dazl protein. While both models are equally possible based on the information currently available, it should be noted that a role for mTORC1 signalling in the post-transcriptional regulation of *Dazl* expression has previously been described in mouse embryonic stem cells (Sampath et al., 2008). As such, I am currently developing a system to robustly differentiate between total and translating mRNA levels in small cell numbers to further investigate the extent of post-transcriptional regulation during mouse PGC development.

While I have provided some insight into the downstream targets of Akt signalling in E10.5 PGCs, the question remains as to how exactly Akt signalling is itself activated. With that in mind, retinoic acid has previously been shown to activate Akt/mTORC1 signalling in mouse embryonic stem cells (Carnevalli et al., 2010), and robust expression of *Aldh1a2*, which encodes the major enzyme of retinoic acid synthesis in the mouse embryo, has previously been shown to be expressed in the genital ridge of both sexes as early as E10.5 (Bowles et al., 2006). Alternatively, steel/kit signalling has also been shown to activate Akt signalling in PGCs *in vitro* (De Miguel et al., 2002), although it should be noted that Steel/kit interaction appears to be ongoing between E7.5 and E11.5 *in vivo* (Gu et al., 2009) while Akt activation only occurs starting at E9.5 (with full activation observed by E10.5). Nonetheless, considering these observations and the results I present above, the relationship between retinoic acid, steel/kit signalling, activation of Akt/mTORC1 kinases and Dazl protein expression warrants further investigation in future studies.

Chapter 6

General Discussion

6.1 On the mechanism of gonadal DNA demethylation in mouse primordial germ cells

As discussed extensively in detail in Chapter 1, mouse primordial germ cells undergo a wave of near complete DNA demethylation following their colonisation of the genital ridge (Guibert et al., 2012; Hajkova et al., 2002; Lee et al., 2002; Seisenberger et al., 2012). Ultimately, this results in loss of 5mC from normally methylated imprint control regions, germ line gene promoters and repetitive elements (Guibert et al., 2012; Hajkova et al., 2002; Lee et al., 2002; Seisenberger et al., 2012). While it is now well-established that gonadal DNA demethylation in mammalian germ cells represents a critical event for the generation of murine gametes capable of supporting embryonic development in the next generation (see Chapter 1 for details), how this process occurs at a molecular level remains the subject of intense debate (Hill et al., 2014). As such, at the onset of this PhD, my aim was to gain insight into the mechanisms underpinning DNA demethylation in mouse PGCs.

The currently favoured model for gonadal DNA demethylation in PGCs proposes that the 5mC-oxygenase Tet1 critically catalyses the oxidation of 5mC to 5hmC which is subsequently diluted out over rounds of DNA replication and cell division, ultimately resulting in a demethylated PGC population (Hackett et al., 2013; Yamaguchi et al., 2013b). However, early on in my research, I recognised that this model had yet to be functionally validated. Rather, this model had been inferred based on: 1) 5hmC dynamics during reprogramming in wild type PGCs at a few loci (Hackett et al., 2013); and 2) increased total DNA modification levels in late stage (E13.5) *Tet1*-KO PGCs and adult germ cells, as well as in the embryonic offspring of *Tet1*-KO males crossed with wild type females (Yamaguchi et al., 2012; Yamaguchi et al., 2013b). I thus set out to functionally validate this model. My results clearly demonstrate that: 1) DNA demethylation is not impaired in the absence of Tet1; 2) that 5hmC localisation is most consistent with a model whereby 5mC oxidation protects the hypo-methylated state following gonadal DNA demethylation; and 3) that loss of Tet1 results in aberrant removal of residual methylation and/or *de novo* methylation following DNA demethylation at a very

small number of loci. While my proposed model whereby Tet1/5mC oxidation protects the hypo-methylated state following the major loss of 5mC in PGCs represents a significant paradigm shift from the currently accepted model for the role of Tet1 during gonadal DNA demethylation, I stress that my results are fully consistent with previously published results (Hackett et al., 2013; Yamaguchi et al., 2012; Yamaguchi et al., 2013b). My observations regarding the role of Tet1 in genome-wide DNA demethylation in gonadal PGCs are additionally reminiscent of a recent study from our lab detailing the role of Tet3-driven 5mC oxidation during zygotic DNA demethylation (Amouroux et al., 2016). In the mouse zygote, the initial wave of DNA demethylation does not require 5hmC formation; rather, the main role for Tet3 is to protect against zygotic maintenance and *de novo* DNA methylation driven by Dnmt1 and Dnmt3a (Amouroux et al., 2016). In this context, I note that, although globally hypo-methylated, gonadal PGCs also express *Dnmt1* and *Dnmt3a* methyltransferases (Figures S6C-D). Collectively, my work and the study from Amouroux et al. demonstrate the conserved mechanistic role of Tet proteins during epigenetic reprogramming and suggest that global reprogramming events require efficient protection from *de novo* DNA methylation following removal of 5mC to stabilise the newly acquired epigenetic state. It should be noted, however, that the overall biological significance of these two-tier demethylation processes remains to be established. Nonetheless, I also note that, although this specific role of Tet enzymes might be more apparent during reprogramming-associated global DNA demethylation, I envisage a similar role for these proteins in other instances, including in mESCs (Hon et al., 2014; Williams et al., 2011) and in differentiated somatic cells (Jin et al., 2014; Wiehle et al., 2015) .

While the work I describe above provides mechanistic insight into how DNA hypo-methylation is maintained in PGCs following loss of 5mC, unfortunately, beyond dismissing the currently accepted model for gonadal DNA demethylation, it does not offer much further insight into the exact molecular mechanism by which DNA demethylation occurs. Currently, both passive and active models have been proposed (reviewed in (Hill et al., 2014)). Passive DNA demethylation has been inferred based on low expression of factors implicated in DNA methylation (Kagiwada et al., 2013; Seisenberger et

al., 2012; Seki et al., 2005), a lack of clear Dnmt1/PCNA co-localisation during S-phase (Kagiwada et al., 2013), and gradual loss of total DNA modification levels over time (Kagiwada et al., 2013). On the other hand, active DNA demethylation has been inferred based on increased activation of the Base Excision Repair (BER) pathway in PGCs relative to surrounding somatic cells during gonadal epigenetic reprogramming (Hajkova et al., 2010). It should be noted however that, to date, neither passive nor active models for gonadal DNA demethylation in PGCs are supported by results from which definitive conclusions can be drawn. With respect to passive DNA demethylation, although down-regulated, the DNA methyltransferase machinery is clearly expressed in PGCs (except for *Dnmt3l*) and thus may still be active. Moreover, gradual loss of total DNA modification levels is not necessarily inconsistent with an active mechanism of DNA demethylation, especially if PGCs are heterogeneous in the timing of their developmental progression. With respect to active DNA demethylation, it is possible that the activation of the BER machinery at the time of 5mC loss simply reflects potentially high levels of Tet1/Tdg-mediated maintenance DNA demethylation (He et al., 2011) relative to surrounding somatic cells following the major loss of 5mC.

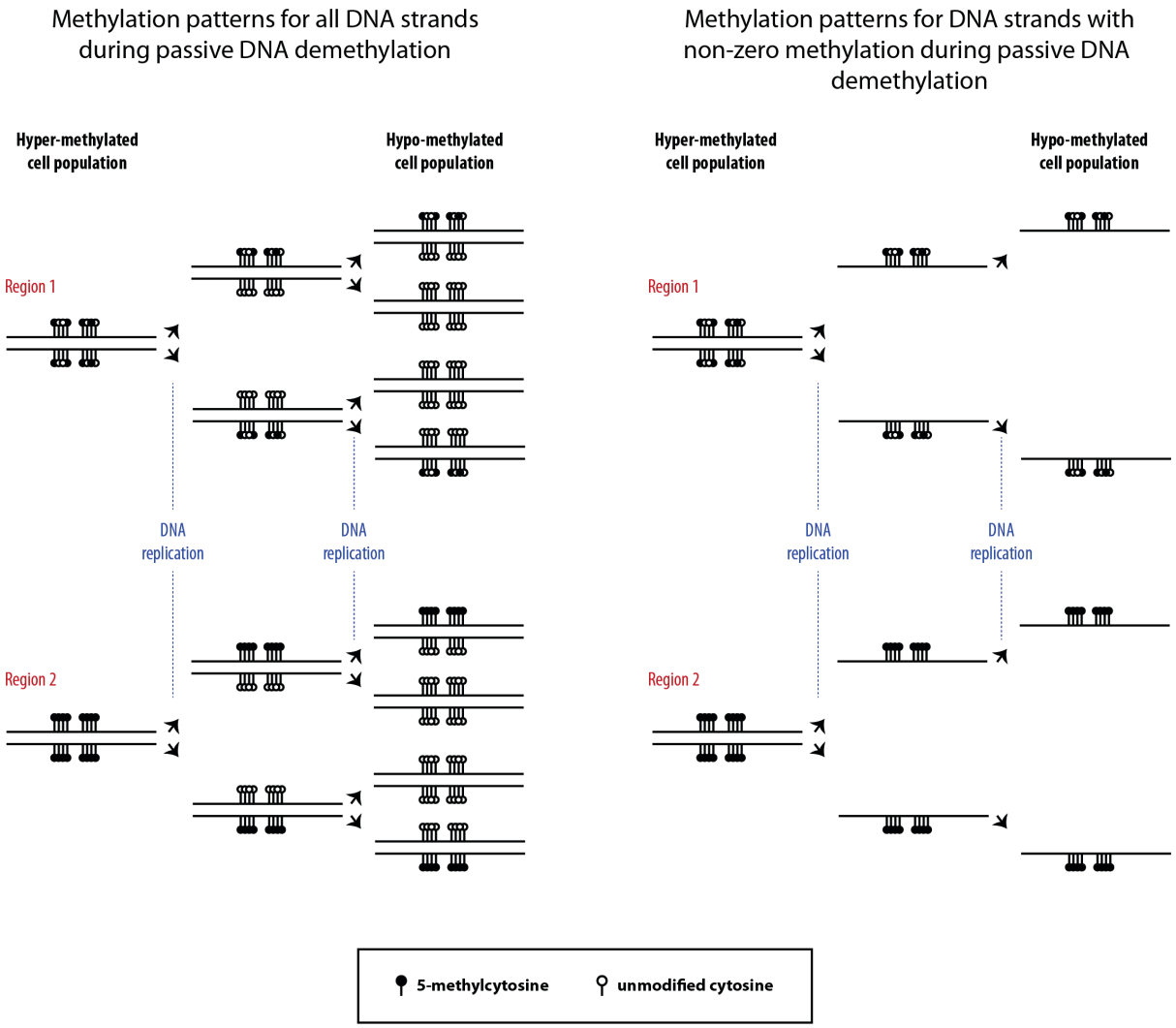


Figure 6.1 – Prediction. *Passive DNA demethylation predicts conservation of methylation entropy at a given genomic locus for all DNA strands with non-zero methylation between hyper-methylated and hypo-methylated cell populations*

Mechanistically, it should be possible in future studies to distinguish between active and passive models of DNA demethylation in two ways. First, as passive DNA demethylation predicts the conservation of methylation patterns over time, it consequently also predicts the conservation of methylation entropy for non-zero methylated DNA strands during gonadal DNA demethylation (Figure 6.1). That is to say, following replication of each double-stranded DNA molecule, the methylation patterns of the two parent strands are retained while the nascent daughter strands are completely unmethylated. Thus, over the course of multiple cell divisions, the methylation patterns of all DNA strands with non-zero methylation should be identical from day-to-day within a population

of cells regardless of the absolute methylation level observed for the population (Figure 6.1). From a mathematical point of view, this should manifest itself as conservation of the methylation entropy for non-zero methylated DNA strands between pre-demethylated and post-demethylated PGCs. As this can be determined from high coverage time-series whole genome bisulphite sequencing experiments (He et al., 2013), this represents a highly interesting and unexplored area of future study.

Second, it should be possible to distinguish between active and passive models of DNA demethylation by monitoring the kinetics of DNA demethylation in PGCs following treatment with aphidicolin, a selective inhibitor of S phase. To date, this has been done for the *H19* locus, in which clear DNA demethylation of the paternally methylated allele is observed in the absence of DNA replication (Kawasaki et al., 2014), suggesting that an active mechanism is driving demethylation at this particular locus. Future studies coupling urogenital ridge culture with aphidicolin treatment followed by LC/MS analysis and genome-wide bisulphite sequencing analysis will be useful to expand on these initial observations.

6.2 On the timing and molecular triggers of gonadal DNA demethylation and licensing for gametogenesis in mouse PGCs

Although loss of Tet1 was originally hypothesized to result in delayed/impaired DNA demethylation in PGCs (Hackett et al., 2013; Yamaguchi et al., 2012; Yamaguchi et al., 2013b), my results clearly suggest that this process is in fact precocious. Consistently, up until E12.5, PGCs isolated from *Tet1*-KO embryos also appear to be more developmentally advanced at the transcriptional level compared to their wild type equivalents. At the molecular level, it remains unclear exactly what is driving this phenotype. One possibility that I discuss in Chapter 4 is that a conserved gene set de-regulated in Tet1-KO PGCs and mESCs (i.e. cluster 1 in Figure 4.6) is enriched for mitotic cell cycle and DNA

repair genes. As these genes are consistently higher expressed, it is possible that potentially higher proliferation rates and/or DNA repair activity may be responsible for increased DNA demethylation via passive and active mechanisms, respectively. However, I emphasize that this remains hypothesis and requires validation.

Alternatively, a second possibility is that precocious development of pre-E12.5 PGCs in *Tet1*-KO embryos is a consequence of de-regulated interaction between germ cells and surrounding somatic cells in the mid-gestation embryo. In this context, it is important to note that I used a *Tet1*-knockout model which results in global deletion of Tet1 protein in all tissue, including soma (Dawlaty et al., 2011). While, it remains to be determined whether gonadal DNA demethylation and licensing for gametogenesis are induced in response to PGCs migrating into the genital ridge environment or whether these processes occur in a cell-autonomous way, evidence to date suggests that response to signalling may play an important part in triggering both of these processes. Specifically, during the *in vitro* derivation of PGC-like cells (PGCLCs), induction of epigenetic reprogramming (including erasure of methylation imprints) and high mRNA expression of genes involved in licensing for gametogenesis (including *Dazl*) requires re-aggregation of PGCLCs with disaggregated somatic cells of the E12.5 female genital ridge (Hayashi et al., 2012). Moreover, the results I present in Chapter 5 expand on these published observations, showing that expression of Dazl protein can be impaired by inhibition of Akt-mTORC1 signalling, a pathway which is specifically activated in PGCs immediately following their colonisation of the genital ridge. Thus, although future co-culture experiments re-aggregating migratory PGCs isolated from E8.5/E9.5 embryos with somatic cells isolated either from the genital ridge or from other tissue in the E11.5 embryo will be necessary to provide a definitive answer as to whether gonadal DNA demethylation and licensing for gametogenesis are induced in response to changes in the PGC micro-environment, it remains entirely possible that de-regulated gene expression in somatic cells surrounding PGCs in *Tet1*-KO embryos ultimately drives precocious PGC development. Similarly, de-regulated interaction between germ cells and surrounding somatic cells in the female gonad may equally be responsible for the impaired meiotic phenotype observed in

later germ cells of *Tet1*-KO embryos. With this is mind, a conditional *Tet1*-KO system may prove invaluable in future studies to decipher the relative contribution of soma and germ cells in the de-regulated timing of PGCs isolated from *Tet1*-KO embryos.

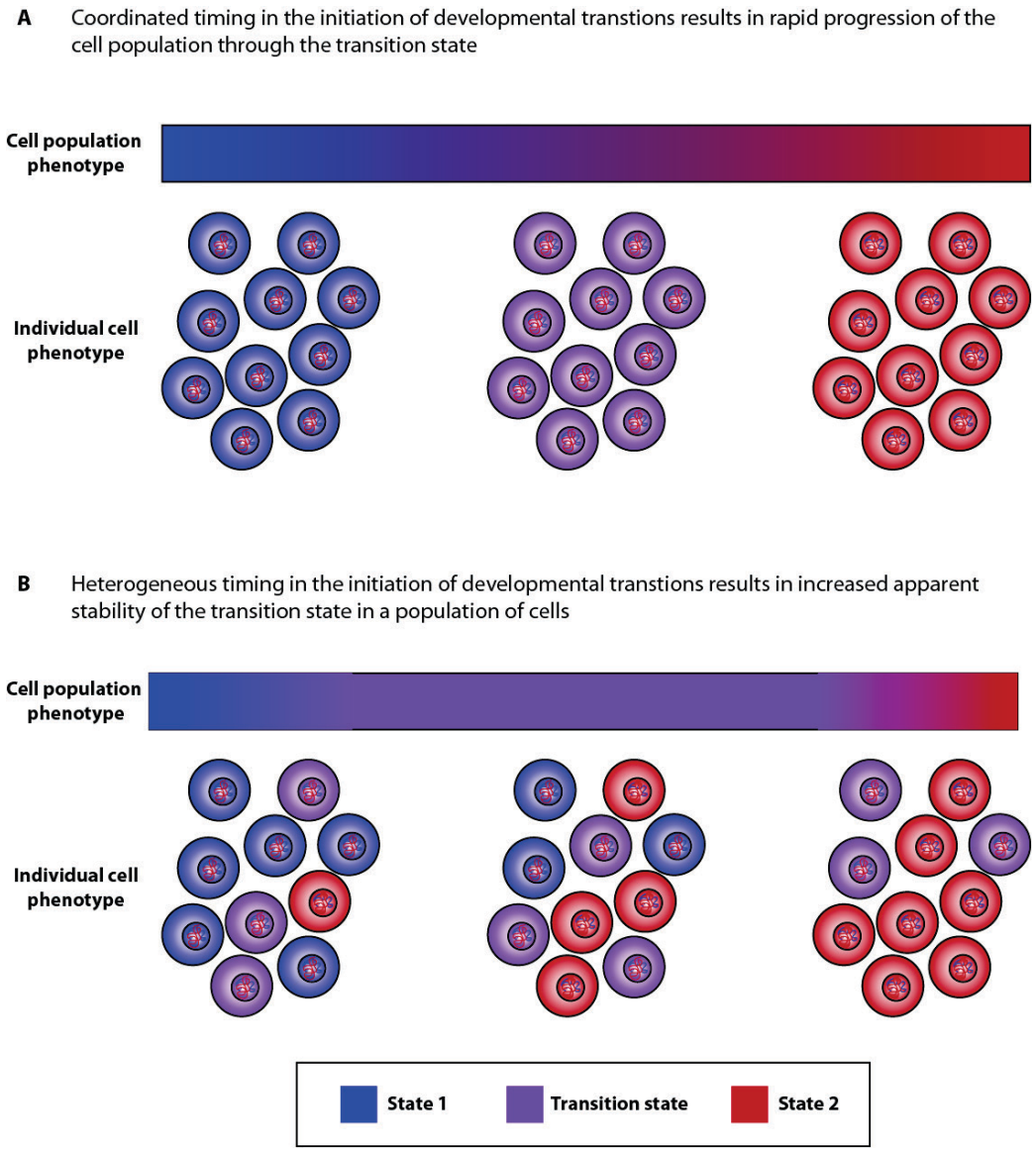


Figure 6.2 – Hypothesis for why Tet1-KO PGCs show increased stability of the demethylated pre-meiotic germ cell state. A) *Coordinated timing in the initiation of a developmental transition results in rapid progression of a population of cells through the transition state.* ***B)*** *Heterogeneity with respect to timing of the initiation of a developmental transition results in increased apparent stability of the transition state at the population level.*

In addition to precocious development of PGCs in pre-E12.5 embryos, I also observed delayed progression of germ cells in post-E12.5 embryos devoid of Tet1. Consequently, as a population, *Tet1*-

KO germ cells appear to have increased stability of the demethylated pre-meiotic germ cell sub-state. It should be noted, however, that stability of a particular developmental/transcriptional sub-state with respect to a *population* of cells is not necessarily a reflection of stability of that particular sub-state with respect to *individual* cells within that population. That is to say, a population of cells transitioning between two transcriptional/developmental states can show increased stability of an intermediate sub-state under one condition (population 1) compared to a second condition (population 2) if individual cells within population 1 show greater heterogeneity with respect to the timing they initiate the transition than individual cells within population 2 (Figure 6.2). In the context of PGCs, wild type germ cells transition from a methylated mitotic state (E10.5) to a de-methylated meiotic (E14.5 female) or de-methylated mitotically-arrested (E14.5 male) state. Loss of Tet1 appears to result in increased stability of a state that is roughly equivalent to an intermediate between E12.5 male and female PGCs and E14.5 male germ cells. It is thus conceivable that the increased stability of this sub-state in *Tet1*-KO PGCs is due to increased heterogeneity in the timing with which germ cells initiate the E10.5 to E14.5 transition. At the transcriptional level, this would appear as increased variability in gene expression of individual *Tet1*-KO PGCs compared with wild type PGCs at a given stage between E10.5 and E14.5. In this context, it is interesting to note that increased variability in gene expression patterns for *Tet1*/*Tet3*-double knockout mouse blastocysts (relative to wild type blastocysts) has recently been described (Kang et al., 2015). Alternatively, it is also conceivable that the apparent increased stability of the demethylated pre-meiotic germ cell state reflects a more synchronous process in PGC development, due to consistently higher expression of pluripotency transcription factors in early PGCs followed by retention of pluripotency transcription factors in late PGCs. Altogether, these observations suggest that single cell transcriptional and methylation analyses of PGCs isolated from *Tet1*-KO embryos will be a highly interesting area of future research.

Bibliography

Adams, I., and McLaren, A. (2002). Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis. Development *129*, 1155-1164.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E.*, et al.* (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol *18*, 225-233.

Ancelin, K., Lange, U., Hajkova, P., Schneider, R., Bannister, A., Kouzarides, T., and Surani, M. (2006). Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. Nat Cell Biol *8*, 623-630.

Anders, S., Pyl, P., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Anderson, E., Baltus, A., Roepers-Gajadien, H., Hassold, T., de Rooij, D., van Pelt, A., and Page, D. (2008). Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. PNAS *105*, 14976-14980.

Anderson, R., Copeland, T., Schöler, H., Heasman, J., and Wylie, C. (2000). The onset of germ cell migration in the mouse embryo. Mech Dev *91*, 61-68.

Ara, T., Nakamura, Y., Egawa, T., Sugiyama, T., Abe, K., Kishimoto, T., Matsui, Y., and Nagasawa, T. (2003). Impaired colonization of the gonads by primordial germ cells in mice lacking a chemokine, stromal cell-derived factor-1 (SDF-1). PNAS *100*, 5319-5323.

Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods *11*, 1138-1140.

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H., Iurlaro, M., Reik, W., Murrell, A., and Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. Nat Chem Biol *11*.

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem *6*, 1049-1055.

Barlow, D., Stöger, R., Herrmann, B., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature *349*, 84-87.

Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F.*, et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

Bartolomei, M., and Ferguson-Smith, A. (2011). Mammalian genomic imprinting. Cold Spring Harb Perspect Biol *3*, pii: a002592.

Bartolomei, M., Webber AL, Brunkow ME, Tilghman SM. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev *7*, 1663-1673.

Best, D., Sahlender, D., Walther, N., Peden, A., and Adams, I. (2008). Sdmg1 is a conserved transmembrane protein associated with germ cell sex determination and germline-soma interactions in mice. Development *135*, 1415-1425.

Blaschke, K., Ebata, K., Karimi, M., Zepeda-Martínez, J., Goyal, P., Mahapatra, S., Tam, A., Laird, D., Hirst, M., Rao, A.*, et al.* (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222-226.

Bock, C., Tomazou, E., Brinkman, A., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H., and Meissner, A. (2010). Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotech *28*, 1106-1114.

Booth, M., Branco, M., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science *336*, 934-937.

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

Bostick, M., Kim, J., Estève, P., Clark, A., Pradhan, S., and Jacobsen, S. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Bourc'his, D., Bestor TH. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

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

Bowles, J., Feng, C., Spiller, C., Davidson, T., Jackson, A., and Koopman, P. (2010). FGF9 suppresses meiosis and promotes male germ cell fate in mice. Dev Cell *19*, 440-449.

Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsaksophak, K., Wilson, M., Rossant, J.*, et al.* (2006). Retinoid signaling determines germ cell fate in mice. Science *2006*, 596-600.

Brook, M., Smith, J., and Gray, N. (2009). The DAZL and PABP families: RNA-binding proteins with interrelated roles in translational control in oocytes. Reproduction *137*, 595-617.

Carnevalli, L., Masuda, K., Frigerio, F., Le Bacquer, O., Um, S., Gandin, V., Topisirovic, I., Sonenberg, N., Thomas, G., and Kozma, S. (2010). S6K1 plays a critical role in early adipocyte differentiation. Dev Cell *18*, 763-774.

Cattanach, B., and Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. Nature *315*, 496-498.

Chang, H., and Matzuk, M. (2001). Smad5 is required for mouse primordial germ cell development. Mech Dev *104*, 61-67.

Chassot, A., Ranc, F., Gregoire, E., Roepers-Gajadien, H., Taketo, M., Camerino, G., de Rooij, D., Schedl, A., and Chaboissier, M. (2008). Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary. Hum Mol Genet *17*, 1264-1277.

Chuang, L., Ian, H., Koh, T., Ng, H., Xu, G., and Li, B. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Chuva de Sousa Lopes, S., Hayashi, K., Shovlin, T., Mifsud, W., Surani, M., and McLaren, A. (2008). X chromosome activity in mouse XX primordial germ cells. PLoS Genet *4*, e30.

Colvin, J., Green, R., Schmahl, J., Capel, B., and Ornitz, D. (2001). Male-to-female sex reversal in mice lacking fibroblast growth factor 9. Cell *104*, 875-889.

Costa, Y., Ding, J., Theunissen, T., Faiola, F., Hore, T., Shliaha, P., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S.*, et al.* (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370-374.

Creyghton, M., Cheng, A., Welstead, G., Kooistra, T., Carey, B., Steine, E., Hanna, J., Lodato, M., Frampton, G., Sharp, P.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. PNAS *2010*, 21931-21936.

Dawlaty, M., Breiling, A., Le, T., Raddatz, G., Barrasa, M., Cheng, A., Gao, Q., Powell, B., Li, Z., Xu, M.*, et al.* (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell *24*, 310-323.

Dawlaty, M., Ganz, K., Powell, B., Hu, Y., Markoulaki, S., Cheng, A., Gao, Q., Kim, J., Choi, S., Page, D.*, et al.* (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell *9*, 166-175.

De Miguel, M., Cheng, L., Holland, E., Federspiel, M., and Donovan, P. (2002). Dissection of the c-Kit signaling pathway in mouse primordial germ cells by retroviral-mediated gene transfer. PNAS *99*, 10458-10463.

de Napoles, M., Nesterova, T., and Brockdorff, N. (2007). Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells. PLoS One *2*, e860.

de Sousa Lopes, S., Roelen, B., Monteiro, R., Emmens, R., Lin, H., Li, E., Lawson, K., and Mummery, C. (2004). BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo. Genes Dev *18*, 1838-1849.

DeChiara, T., Robertson, E., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell *64*, 849-859.

Dokshin, G., Baltus, A., Eppig, J., and Page, D. (2013). Oocyte differentiation is genetically dissociable from meiosis in mice. Nat Genet *45*, 877-883.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat Methods *9*, 215-216.

Ferguson-Smith, A. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet *12*, 565-575.

Ferguson-Smith, A., Cattanach, B., Barton, S., Beechey, C., and Surani, M. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. Nature *351*, 667-670.

Ferguson-Smith, A., Sasaki, H., Cattanach, B., and Surani, M. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature *362*, 751-755.

Ficz, G., Hore, T., Santos, F., Lee, H., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y., Walter, J.*, et al.* (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell *13*, 351-359.

Fu, L., Guerrero, C., Zhong, N., Amato, N., Liu, Y., Liu, S., Cai, Q., Ji, D., Jin, S., Niedernhofer, L.*, et al.* (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc *136*, 11582-11585.

Fujita, P., Rhead, B., Zweig, A., Hinrichs, A., Karolchik, D., Cline, M., Goldman, M., Barber, G., Clawson, H., Coelho, A.*, et al.* (2011). The UCSC Genome Browser database: update 2011. Nucleic Acis Res *39*, D876-882.

Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y.*, et al.* (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell *12*, 453-469.

Gardner, R., Lyon, M., Evans, E., and Burtenshaw, M. (1985). Clonal analysis of X-chromosome inactivation and the origin of the germ line in the mouse embryo. J Embryol Exp Morphol *88*, 349-363.

Gill, M., Hu, Y., Lin, Y., and Page, D. (2011). Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells. PNAS *108*.

Ginsburg, M., Snow, M., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. Development *110*, 521-528.

Gkountela, S., Zhang, K., Shafiq, T., Liao, W., Hargan-Calvopiña, J., Chen, P., and Clark, A. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. Cell *161*, 1425-1436.

Goldberg, A., Allis, C., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Grabole, N., Tischler, J., Hackett, J., Kim, S., Tang, F., Leitch, H., Magnúsdóttir, E., and Surani, M. (2013). Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. EMBO Rep *14*, 629-637.

Grskovic, M., Chaivorapol, C., Gaspar-Maia, A., Li, H., and Ramalho-Santos, M. (2007). Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. PLoS Genet *3*, e145.

Gu, T., Guo, F., Yang, H., Wu, H., Xu, G., Liu, W., Xie, Z., Shi, L., He, X., Jin, S.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Gu, Y., Runyan, C., Shoemaker, A., Surani, A., and Wylie, C. (2009). Steel factor controls primordial germ cell survival and motility from the time of their specification in the allantois, and provides a continuous niche throughout their migration. Development *136*, 1295-1303.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res *22*, 633-641.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T., Hu, B.*, et al.* (2014a). Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell *15*, 447-458.

Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y.*, et al.* (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. Cell *161*, 1437-1452.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J.*, et al.* (2014b). The DNA methylation landscape of human early embryos. Nature *511*, 606-610.

Habibi, E., Brinkman, A., Arand, J., Kroeze, L., Kerstens, H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.*, et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360-369.

Hackett, J., Sengupta, R., Zylicz, J., Murakami, K., Lee, C., Down, T., and Surani, M. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448-452.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature *452*, 877-881.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech Dev *117*, 15-23.

Hajkova, P., Jeffries, S., Lee, C., Miller, N., Jackson, S., and Surani, M. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science *329*, 78-82.

Handel, M., and Schimenti, J. (2010). Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat Rev Genet *11*, 124-136.

Hashimoto, H., Pais, J., Zhang, X., Saleh, L., Fu, Z., Dai, N., Corrêa Jr, I., Zheng, Y., and Cheng, X. (2014). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature *506*, 391-395.

Hassold, T., Hall, H., and Hunt, P. (2007). The origin of human aneuploidy: where we have been, where we are going. Hum Mol Genet *16 Spec No. 2*, R203-208.

Hayashi, K., Ogushi, S., Kurimoto, K., Shimamoto, S., Ohta, H., and Saitou, M. (2012). Offspring from oocytes derived from in vitro primordial germ cell-like cells in mice. Science *338*, 971-975.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519-532.

He, J., Sun, X., Shao, X., Liang, L., and Xie, H. (2013). DMEAS: DNA methylation entropy analysis software. Bioinformatics *29*, 2044-2045.

He, Y., Li, B., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

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

Hirasawa, R., and Sasaki, H. (2009). Dynamic transition of Dnmt3b expression in mouse pre- and early post-implantation embryos. Gene Expr Patterns *9*, 27-30.

Hon, G., Song, C., Du, T., Jin, F., Selvaraj, S., Lee, A., Yen, C., Ye, Z., Mao, S., Wang, B.*, et al.* (2014). 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. Mol Cell *56*, 286-297.

Howell, C., Bestor, T., Ding, F., Latham, K., Mertineit, C., Trasler, J., and Chaillet, J. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell *104*, 829-838.

Hu, X., Zhang, L., Mao, S., Li, Z., Chen, J., Zhang, R., Wu, H., Gao, J., Guo, F., Liu, W.*, et al.* (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell *14*, 512-522.

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W., Kang, J., Zepeda-Martínez, J., Pape, U., Jacobsen, S., Peters, B.*, et al.* (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. PNAS *111*, 1361-1366.

Huang, Y., Pastor, W., Shen, Y., Tahiliani, M., Liu, D., and Rao, A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One *5*, e8888.

Hyldig, S., Croxall, N., Contreras, D., Thomsen, P., and Alberio, R. (2011). Epigenetic reprogramming in the porcine germ line. BMC Dev Biol *11*, 11.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.

Ito, S., D'Alessio, A., Taranova, O., Hong, K., Sowers, L., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R.*, et al.* (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. Nat Genet *46*, 17-23.

Jin, C., Lu, Y., Jelinek, J., Liang, S., Estecio, M., Barton, M., and Issa, J. (2014). TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. Nucleic Acis Res *42*, 6956-6971.

Jurka, J., Kapitonov, V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res *110*, 462-467.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J *32*, 340-353.

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

Kaneko, S., Son, J., Shen, S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat Struct Mol Biol *20*, 1258-1264.

Kang, J., Lienhard, M., Pastor, W., Chawla, A., Novotny, M., Tsagaratou, A., Lasken, R., Thompson, E., Surani, M., Koralov, S.*, et al.* (2015). Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. PNAS *112*, E4236-4245.

Karimi, M., Goyal, P., Maksakova, I., Bilenky, M., Leung, D., Tang, J., Shinkai, Y., Mager, D., Jones, S., Hirst, M.*, et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell *8*, 676-687.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007a). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet *16*, 2272-2280.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007b). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet *2007*, 2272-2280.

Kawasaki, Y., Lee, J., Matsuzawa, A., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2014). Active DNA demethylation is required for complete imprint erasure in primordial germ cells. Sci Rep *4*, 3658.

Kellinger, M., Song, C., Chong, J., Lu, X., He, C., and Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol *19*, 831-833.

Kelly, S. (1977). Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. J Exp Zool *200*, 365-376.

Kim, Y., Kobayashi, A., Sekido, R., DiNapoli, L., Brennan, J., Chaboissier, M., Poulat, F., Behringer, R., Lovell-Badge, R., and Capel, B. (2006). Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination. PLoS Biol *4*, e187.

Kocer, A., Reichmann, J., Best, D., and Adams, I. (2009). Germ cell sex determination in mammals. Hum Mol Reprod *15*, 205-213.

Kono, T., Obata, Y., Wu, Q., Niwa, K., Ono, Y., Yamamoto, Y., Park, E., Seo, J., and Ogawa, H. (2004). Birth of parthenogenetic mice that can develop to adulthood. Nature *428*, 860-864.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. Nature *351*, 117-121.

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

Krueger, F., and Andrews, S. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Kumar, S., Chatzi, C., Brade, T., Cunningham, T., Zhao, X., and Duester, G. (2011). Sex-specific timing of meiotic initiation is regulated by Cyp26b1 independent of retinoic acid signalling. Nat Commun *2*, 151.

Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., Moritoki, Y., Kohri, K., Kimura, H., Yamamoto, T.*, et al.* (2015). Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. Cell Stem Cell *16*, 517-532.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev *22*, 1617-1635.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lawson, K., Dunn, N., Roelen, B., Zeinstra, L., Davis, A., Wright, C., Korving, J., and Hogan, B. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev *13*.

Lawson, K., and Hage, W. (1994). Clonal analysis of the origin of primordial germ cells in the mouse. Ciba Found Symp *182*.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development *129*, 1807-1817.

Leitch, H., Blair, K., Mansfield, W., Ayetey, H., Humphreys, P., Nichols, J., Surani, M., and Smith, A. (2010). Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. Development *137*, 2279-2287.

Leitch, H., McEwen, K., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J., Smith, A.*, et al.* (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol *20*, 311-316.

Lesch, B., Dokshin, G., Young, R., McCarrey, J., and Page, D. (2013). A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. PNAS *110*, 16061-16066.

Lesch, B., and Page, D. (2012). Genetics of germ cell development. Nat Rev Genet *13*, 781-794.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-365.

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

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell *15*, 547-557.

Lin, Y., Gill ME, Koubova J, Page DC. (2008). Germ cell-intrinsic and -extrinsic factors govern meiotic initiation in mouse embryos. Science *322*, 1685-1687.

Lin, Y., and Page, D. (2005). Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. Dev Biol *288*, 309-316.

Lorsbach, R., Moore, Mathew, S., Raimondi, S., Mukatira, S., and Downing, J. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia *17*, 637-641.

Love, M., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev *28*, 2103-2119.

Maatouk, D., Kellam, L., Mann, M., Lei, H., Li, E., Bartolomei, M., and Resnick, J. (2006). DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. Development *133*, 3411-3418.

MacLean, G., Li, H., Metzger, D., Chambon, P., and Petkovich, M. (2007). Apoptotic extinction of germ cells in testes of Cyp26b1 knockout mice. Endocrinology *148*, 4560-4567.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat Cell Biol *15*, 905-915.

Manning, B., and Cantley, L. (2007). AKT/PKB signaling: navigating downstream. Cell *129*, 1261-1274.

Matsui, Y., Zsebo, K., and Hogan, B. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell *37*, 179-183.

McLaren, A. (2003). Primordial germ cells in the mouse. Dev Biol *262*, 1-15.

McLaren, A., and Lawson, K. (2005). How is the mouse germ-cell lineage established? Differentiation *73*, 435-437.

McLaren, A., and Southee, D. (1997). Entry of mouse embryonic germ cells into meiosis. Dev Biol *187*, 107-113.

McLean, C., Bristor, D., Hiller, M., Clarke, S., Schaar, B., Lowe, C., Wenger, A., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotech *28*.

Meissner, A., Gnirke, A., Bell, G., Ramsahoye, B., Lander, E., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acis Res *33*, 5868-5877.

Miyazono, K., and Miyazawa, K. (2002). Id: a target of BMP signaling. Sci STKE *151*, e40.

Moe-Behrens, G., Klinger, F., Eskild, W., Grotmol, T., Haugen, T., and De Felici, M. (2003). Akt/PTEN signaling mediates estrogen-dependent proliferation of primordial germ cells in vitro. Mol Endocrinol *17*, 2630-2638.

Molyneaux, K., Zinszner, H., Kunwar, P., Schaible, K., Stebler, J., Sunshine, M., O'Brien, W., Raz, E., Littman, D., Wylie, C*., et al.* (2003). The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. 130.

Morales-Ruiz, T., Ortega-Galisteo, A., Ponferrada-Marín, M., Martínez-Macías, M., Ariza, R., and Roldán-Arjona, T. (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. PNAS *103*, 6853-6858.

Nabel, C., Jia, H., Ye, Y., Shen, L., Goldschmidt, H., Stivers, J., Zhang, Y., and Kohli, R. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol *8*, 751-758.

Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y., and Saitou, M. (2013). Induction of mouse germ-cell fate by transcription factors in vitro. Nature *501*, 222-226.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M.*, et al.* (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol *9*, 64-71.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Dal Bello, F., and Oliviero, S. (2015). Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. Cell Rep *pii: S2211-1247*, 00009-00001.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. (2013). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. Genome Biol *14*, R91.

Ng, J., Kumar, V., Muratani, M., Kraus, P., Yeo, J., Yaw, L., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. Dev Cell *24*, 324-333.

Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development *136*, 3215-3222.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. Cell *137*, 571-584.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A.*, et al.* (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207-213.

Okamoto, Y., Yoshida, N., Suzuki, T., Shimozawa, N., Asami, M., Matsuda, T., Kojima, N., Perry, A., and Takada, T. (2016). DNA methylation dynamics in mouse preimplantation embryos revealed by mass spectrometry. Sci Rep *6*, 19134.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Res *62*, 4075-4080.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol *10*, 475-478.

Paland, S., and Lynch, M. (2006). Transitions to asexuality result in excess amino acid substitutions. Science *311*, 990-992.

Pastor, W., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol *14*, 341-356.

Pastor, W., Pape, U., Huang, Y., Henderson, H., Lister, R., Ko, M., McLoughlin, E., Brudno, Y., Mahapatra, S., Kapranov, P.*, et al.* (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394-397.

Peat, J., Dean, W., Clark, S., Krueger, F., Smallwood, S., Ficz, G., Kim, J., Marioni, J., Hore, T., and Reik, W. (2014). Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. Cell Rep *9*, 1990-2000.

Popp, C., Dean, W., Feng, S., Cokus, S., Andrews, S., Pellegrini, M., Jacobsen, S., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Quinlan, A., and Hall, I. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ramsköld, D., Luo, S., Wang, Y., Li, R., Deng, Q., Faridani, O., Daniels, G., Khrebtukova, I., Loring, J., Laurent, L.*, et al.* (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotech *30*, 777-782.

Reik, W., Collick, A., Norris, M., Barton, S., and Surani, M. (1987). Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature *328*, 248-251.

Robertson, A., Dahl, J., vågbø, C., Tripathi, P., Krokan, H., and Klungland, A. (2011). A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acis Res *39*, e55.

Ruggiu, M., Speed, R., Taggart, M., McKay, S., Kilanowski, F., Saunders, P., Dorin, J., and Cooke, H. (1997). The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. Nature *389*, 73-77.

Sachs, M., Onodera, C., Blaschke, K., Ebata, K., Song, J., and Ramalho-Santos, M. (2013). Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell Rep *2*, 1777-1784.

Saitou, M., Barton, S., and Surani, M. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293-300.

Sampath, P., Pritchard, D., Pabon, L., Reinecke, H., Schwartz, S., Morris, D., and Murry, C. (2008). A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. Cell Stem Cell *2*, 448-460.

Saunders, P., Turner, J., Ruggiu, M., Taggart, M., Burgoyne, P., Elliott, D., and Cooke, H. (2003). Absence of mDazl produces a final block on germ cell development at meiosis. Reproduction *126*, 589-597.

Saxena, R., Brown, L., Hawkins, T., Alagappan, R., Skaletsky, H., Reeve, M., Reijo, R., Rozen, S., Dinulos, M., Disteche, C.*, et al.* (1996). The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat Genet *14*, 292-299.

Schöler, H., Dressler, G., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO J *9*, 2185-2195.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell *48*, 849-862.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol *278*, 440-458.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development *134*, 2627-2638.

Sekido, R., Bar, I., Narváez, V., Penny, G., and Lovell-Badge, R. (2004). SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. Dev Biol *274*, 271-279.

Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature *453*, 930-934.

Seligman, J., and Page, D. (1998). The Dazh gene is expressed in male and female embryonic gonads before germ cell sex differentiation. Biochem Biophys Res Commun *245*, 878-882.

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

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. Cell Stem Cell *15*, 459-470.

Shi, F., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M., Wan, M., and Songyang, Z. (2013). Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. J Biol Chem *288*, 20776-20784.

Shovlin, T., Durcova-Hills, G., Surani, A., and McLaren, A. (2008). Heterogeneity in imprinted methylation patterns of pluripotent embryonic germ cells derived from pre-migratory mouse germ cells. Dev Biol *313*, 674-681.

Smallwood, S., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet *43*, 811-814.

Smith, Z., Chan, M., Humm, K., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611-615.

Song, C., Szulwach, K., Dai, Q., Fu, Y., Mao, S., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H.*, et al.* (2013a). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678-691.

Song, C., Szulwach, K., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C., Zhang, W., Jian, X.*, et al.* (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotech *29*, 68-72.

Song, Q., Decato, B., Hong, E., Zhou, M., Fang, F., Qu, J., Garvin, T., Kessler, M., Zhou, J., and Smith, A. (2013b). A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS One *8*, e81148.

Stöger, R., Kubicka, P., Liu, C., Kafri, T., Razin, A., Cedar, H., and Barlow, D. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. Cell *73*, 61-71.

Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., Paulovich, A., Pomeroy, S., Golub, T., Lander, E.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. PNAS *102*, 15545-15550.

Sugimoto, M., and Abe, K. (2007). X chromosome reactivation initiates in nascent primordial germ cells in mice. PLoS Genet *3*, e116.

Sun, Z., Terragni, J., Borgaro, J., Liu, Y., Yu, L., Guan, S., Wang, H., Sun, D., Cheng, X., Zhu, Z.*, et al.* (2013). High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. Cell Rep *3*, 567-576.

Surani, M., Barton, S., and Norris, M. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature *308*, 548-550.

Suzuki, A., and Saga, Y. (2008). Nanos2 suppresses meiosis and promotes male germ cell differentiation. Genes Dev *22*, 430-435.

Tada, T., Tada, M., Hilton, K., Barton, S., Sado, T., Takagi, N., and Surani, M. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. Dev Genes Evol *207*, 551-561.

Tahiliani, M., Koh, K., Shen, Y., Pastor, W., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L., Liu, D., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

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

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W.*, et al.* (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254-1269.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell *6*, 468-478.

Tang, W., Dietmann, S., Irie, N., Leitch, H., Floros, V., Badshaw, C., Hackett, J., Chinnery, P., and Surani, M. (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell *161*, 1453-1467.

Tilghman, S. (1999). The sins of the fathers and mothers: genomic imprinting in mammalian development. Cell *96*, 185-193.

Torres-Padilla, M., Bannister, A., Hurd, P., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int J Dev Biol *50*, 455-461.

Trapnell, C., Pachter, L., and Salzberg, S. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Tremblay, K., Dunn, N., and Robertson, E. (2001). Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development *128*, 3609-3621.

Tsuda, M., Sasaoka, Y., Kiso, M., Abe, K., Haraguchi, S., Kobayashi, S., and Saga, Y. (2003). Conserved role of nanos proteins in germ cell development. Science *301*, 1239-1241.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.*, et al.* (2006). Maintenance of self-renewal ability of mouse

embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*.

Tzivion, G., Dobson, M., and Ramakrishnan, G. (2011). FoxO transcription factors; Regulation by AKT and 14-3-3 proteins. Biochim Biophys Acta *13*, 1938-1945.

Vainio, S., Heikkilä, M., Kispert, A., Chin, N., and McMahon, A. (1999). Female development in mammals is regulated by Wnt-4 signalling. Nature *397*, 405-409.

Valinluck, V., and Sowers, L. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res *67*, 946-950.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K.*, et al.* (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell *49*, 645-656.

Vincent, J., Huang, Y., Chen, P., Feng, S., Calvopiña, J., Nee, K., Lee, S., Le, T., Yoon, A., Faull, K.*, et al.* (2013). Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. Cell Stem Cell *12*, 470-478.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W.*, et al.* (2014). Programming and inheritance of parental DNA methylomes in mammals. Cell *157*, 979-991.

Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X.*, et al.* (2015). Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature *523*, 621-625.

Weber, M., Hellmann, I., Stadler, M., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Weber, S., Eckert, D., Nettersheim, D., Gillis, A., Schäfer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.*, et al.* (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biol Rep *82*, 214-223.

Wiehle, L., Raddatz, G., Musch, T., Dawlaty, M., Jaenisch, R., Lyko, F., and Breiling, A. (2015). Tet1 and Tet2 protect DNA methylation canyons against hypermethylation. Mol Cell Biol, pii: MCB.00587-00515. [Epub ahead of print].

Williams, K., Christensen, J., Pedersen, M., Johansen, J., Cloos, P., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun *2*, 241.

Wu, H., D'Alessio, A., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.

Yamaguchi, H., and Wang, H. (2001). The protein kinase PKB/Akt regulates cell survival and apoptosis by inhibiting Bax conformational change. Oncogene *20*, 7779-7786.

Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K., and Zhang, Y. (2013a). Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res *23*, 329-339.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443-447.

Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2005). Nanog expression in mouse germ cell development. Gene Expr Patterns *5*, 639-646.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013b). Role of Tet1 in erasure of genomic imprinting. Nature *504*, 460-464.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet *40*, 1016-1022.

Yang, X., Zhang, F., and Kudlow, J. (2002). Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. Cell *110*, 69-80.

Yeom, Y., Fuhrmann, G., Ovitt, C., Brehm, A., Ohbo, K., Gross, M., Hübner, K., and Schöler, H. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881-894.

Ying, Q., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-.

Ying, Y., Liu, X., Marble, A., Lawson, K., and Zhao, G. (2000). Requirement of Bmp8b for the generation of primordial germ cells in the mouse. Mol Endocrinol *14*, 1053-1063.

Ying, Y., and Zhao, G. (2001). Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. Dev Biol *232*, 484-492.

Yokobayashi, S., Liang, C., Kohler, H., Nestorov, P., Liu, Z., Vidal, M., van Lohuizen, M., Roloff, T., and Peters, A. (2013). PRC1 coordinates timing of sexual differentiation of female primordial germ cells. Nature *495*, 236-240.

Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Schöler, H., and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Dev Growth Differ *41*, 675-684.

Yu, M., Hon, G., Szulwach, K., Song, C., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B.*, et al.* (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368-1380.

Zhang, Y., Liu, T., Meyer, C., Eeckhoute, J., Johnson, D., Bernstein, B., Nusbaum, C., Myers, R., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.