SPECIFICATION AND EPIGENETIC RESETTING OF THE PIG GERMLINE EXHIBIT CONSERVATION WITH THE HUMAN LINEAGE

by

QIFAN ZHU, BSc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Jan 2021



Division of Animal Sciences

School of Biosciences

The University of Nottingham

DECLARATION

I hereby declare that this thesis is my own work and that it has not been submitted anywhere for any other degree or award. The work presented herein is my own work and where other sources of information have been used, they have been duly acknowledged.

Qifan Zhu

Abstract

Primordial germ cells (PGCs) are the founder cells of germline which follow a specific programme of development characterized by a unique transcriptional network and the resetting of epigenome. Studies in human PGCs (hPGCs) are limited to late migratory and gonadal stages as earlier embryos are almost not accessible. Previous study showed that pig PGC (pPGC) share similar developmental programme to human germ cells in vitro. Therefore, pig may serve as an alternative model to provide insight on transcriptional network and epigenetic reprogramming applicable to hPGC. Using single-cell RNA-seq and whole-genome bisulfite-seq with post-bisulfite adaptor tagging (PBAT-seq), here I show that extensive epigenetic reprogramming, including active DNA demethylation, depletion of macroH2A1, global enrichment of H3K27me3 and X chromosome reactivation (XCR), starts in pre- and early migratory pPGC. Concomitantly, there is dampening of glycolytic gene expression and re-expression of some pluripotency genes like those in pre-implantation embryos. Evolutionarily young transposable elements (TEs) and gene coding regions resistant to DNA demethylation, i.e. escapees, have also been identified in extremely hypomethylated gonadal pPGCs. Some of those aforementioned events show conservation with hPGCs but not mPGCs, suggesting that the pig is a relevant model for studies on non-rodent PGC specification and transgenerational epigenetic inheritance. Detailed insights into the pig germline

will likely contribute to advances in human germline biology, including in vitro gametogenesis.

Acknowledgement

Firstly, I would like to express my sincere appreciation to my supervisor Prof. Ramiro Alberio for his continuous support and immense patience during my PhD study. Without his guidance and constant feedback this PhD would not have been achievable. I would also like to thank my second supervisor Dr Reinhard Stöger and internal accessor Prof Kevin Sinclair for their valuable advice and questions which encourage me to think from different perspectives. Many thanks to the bioinformaticians Fei Sang and Sabine Dietmann. Without their precious guidance and support it would not be possible to complete this project. I'm also extremely grateful to Azim Surani's group, especially Walfred Tang, for providing the hPGC samples, the PBAT protocol and valuable comments. My sincere thanks also go to Petra Hajkova's group for carrying out the LC/MS-MS experiments with our samples.

I wish to express my deepest gratitude to our group members Doris Klisch, Sarah Withey and Priscila Ramos for their encouragement, guidance and valuable input. I'm also very grateful to our technicians Kamila Derecka, Lydia Kwong, Ralph Hourd and Leigh Silvester, who helped me in numerous ways during various stages of my PhD. Also, thanks go to my fellow PhD students especially Connie Clare for making the office a friendly, joyful and productive workplace.

Special thanks to my family and friends for all the unconditionally support and encouragement.

I also gratefully acknowledge the funding for my PhD study received from University of Nottingham and China Scholarship Council.

Contents

Abbreviation.....10

Chapter 1 Introduction.....13

1.1 Ontogeny and Development of Primordial Germ Cell.....13

1.1.1 Development of mammalian PGC.....13

1.1.2 Key difference between rodent and non-rodent PGC development. 23

1.2 Epigenetic reprogramming in mammalian development..... 39

1.2.1 DNA modification during mammalian development..... 40

1.2.2 Histone variants and modification during mammalian development50

1.2.3 X chromosome status during mammalian development.....61

1.2.4 ncRNA during mammalian development..... 65

1.2.5 Epigenetic inheritance..... 66

1.3 In Vitro systems for mammalian germline studies..... 69

1.3.1 Generation and characterisation of PGCLCs in different species.... 70

1.3.2 Applications of in vitro system of mammalian PGCLC.....72

1.4 Hypothesis and Aim.....74

Chapter 2 Transcriptional network of pre-/early migratory and gonadal pig PGCs..... 76

2.1 Introduction..... 76

2.2 Material and Method.....76

2.2.1 Embryo and Gonadal tissue collection..... 76

2.2.2 Immunofluorescence.....77

2.2.3 Tissue dissociation and Fluorescence-activated cell sorting (FACS) for PGC collection..... 78

2.2.4 Single-cell library preparation..... 78

2.2.5 Bioinformatic analysis..... 79

2.3 Results..... 83

2.3.1 Single-cell transcriptome of pre-migratory and gonadal pPGC reveal highly dynamic states of PGC development..... 83

2.3.2 Epigenetic reprogramming events occur early in pre-migratory pPGCs..... 119

2.4 Summary..... 136

Chapter 3 DNA Methylome of Gonadal pPGC..... 140

3.1 Introduction..... 140

3.2 Material and Method..... 141

3.2.1 Gonad collection and digestion and FACS for high purity pPGCs 141

3.2.2 PBAT Library Preparation and sequencing..... 142

3.2.3 Bioinformatic analysis..... 143

3.3 Results..... 145

3.3.1 Purity and yield of FACS-sorted Sda/GM2^{+} cells (pPGC)..... 145

3.3.2 Modification of PBAT library Preparation and Quality of the PBAT-seq data..... 149

3.3.3 General description of gonadal pPGC methylome..... 151

3.3.4 Escapee in pPGCs show similar and distinct features compared to other species.....162

3.3.5 Methylation profiles of the TE-poor escapee regions in pig gonadal somatic cells and blastocysts..... 171

3.4 Summary..... 175

Chapter 4 General Discussion and Further Work.....177

4.1 Dual origin of PGC in non-rodent mammals?..... 178

4.2 Unique epigenetic signatures may contribute to pPGC specification and development..... 180

4.2.1 5hmC enrichment in newly specified pPGCs..... 180

4.2.2 Absence of macroH2A1 throughout pPGC development and the early enrichment of H3K27me3 in pre-gonadal pPGCs..... 181

4.2.3 Early onset of XCR in pPGCs..... 183

4.2.4 Increased transcription of TE in pPGCs..... 185

4.3 Escapees and transgenerational epigenetic inheritance..... 186

4.4 Summary..... 188

Reference.....189

Appendix.....243

Abbreviation

5caC, 5-carboxylcytosine

5fC, 5-formylcytosine

5hmC, 5-hydroxymethylcytosine

5mC, 5-methylcytosine

αKG, α-ketoglutarate

acetyl-CoA, acetyl coenzyme A

Al, allantois

AVE, anterior visceral endoderm

BER, base excision repair

$BVSC^{+}$, Blimp1-mVenus and Stella-ECFP positive

CGIs, CpG islands

DE, distal endoderm

DEG, differentially expressed gene

DSB, double-strank break

DVE, distal visceral endoderm

DNMT, DNA methyltransferase

ECs, embryonal carcinomas

EM, embryonic mesoderm

EMT, Epithelial–mesenchymal transition

Epi, epiblast

EpiLC, epiblast-like cell

EPSC, expanded potential stem cell

ETC, electron transport chain

ExE, extra-embryonic ectoderm

ExM, extra-embryonic mesoderm

FA, Fanconi anaemia

FACS, Fluorescence-activated cell sorting

GO, Gene Ontology

H3K9me2/3, di-/tri-methylation of lysine 9 on histone H3

H3K27me3, tri-methylation of lysine 27 on histone H3

HyperMR, hypermethylated region

IAP, intracisternal A particle

Abbreviation

ICLs, interstrand cross-links
IF, immunofluorescence
iMeLC, incipient mesoderm-like cell
iPSC, induced pluripotent stem cells
LC-MS/MS, liquid chromatography–tandem mass spectrometry
miRNA, micro-RNA
mtDNA, mitochondrial DNA
OxPhos, oxidative phosphorylation
PBAT-seq, whole genome bisulfite sequencing with post-bisulfite adapter tagging
PGCs, primordial germ cells
PGCLCs, primordial germ cell-like cells
piRNA, PIWI-interacting RNAs
PSC, pluripotent stem cell
RA, retinoic acid
RPM, Reads Per Kilobase Million
SAM, S-adenosylmethionine
Sm, somite
sncRNA, small non-coding RNA
SNP, single nucleotide polymorphism
SNV, single nucleotide variation
t-SNE, t-stochastic neighbour embedding
TE, transposable element
TEI, transgenerational epigenetic inheritance
TET, Ten-eleven translocator
TPM, Transcripts Per Kilobase Million
UNC, unsupervised hierarchal clustering
UMAP, uniform manifold approximation and projection
VE, visceral endoderm.
XCI, X chromosome inactivation
XCR, X chromosome reactivaton
Xi, Inactive X chromosome
XIC, X-chromosome inactivation centre
Xm, Maternal X chromosome

Abbreviation

Xp, Paternal X chromosome

Xr, reactivating X chromosome

Chapter 1 Introduction

1.1 Ontogeny and Development of Primordial Germ Cell

1.1.1 Development of mammalian PGC

The germ line of an organism provides a continuing link between all generations. After the oocyte is fertilized by the sperm, a totipotent zygote is generated and has the ability to develop into a whole organism, including its own germ cells. In mammals, primordial germ cells (PGCs) are the founder cells of the germ line. PGCs are specified at peri-gastrulation in the posterior end of an embryo, and then migrate to the genital ridges through the developing hindgut endoderm. In the developing gonad, PGCs undergo meiosis and differentiate into gametes (Tang et al., 2016, Saitou and Yamaji, 2012).

Studies of PGC development were mainly carried out in mice until recently. Despite the overall similarity between mouse and human PGC development, recent studies of human PGC (hPGC), both *in vivo* and *in vitro*, have reported key differences compared to mouse PGC (mPGC) at signalling, transcriptional and epigenetic levels (Tang et al., 2015, Irie et al., 2015, von Meyenn et al., 2016). On the other hand, early PGC development in pigs shows striking similarity with humans and monkeys *in vitro* induction systems (Kobayashi et al., 2017, Sasaki et al., 2016). In this section, the regulatory network and unique events during mammalian PGC development will be discussed, with a focus on key difference between species.

1.1.1.1 PGC specification and development in the mouse

A small cluster of six mPGC precursors marked by Blimp1 were identified around E6-E6.5 mouse embryos (Saitou et al., 2002). These Blimp1-positive cells repress Hox genes and express *Sox2, Stella, and Nanog* from E6.75 onward (Yamaguchi et al., 2005, Kurimoto et al., 2008). The number of mPGC in the cluster increases to around 40 at E7.25. From around E7.5 onwards, newly specified PGCs initiate migration towards developing genital ridges through definitive hindgut endoderm, during which PGC enter G2 cell cycle

arrest at around E7.75 (Seki et al., 2007). After leaving the back of allantois, mPGCs in the hindgut first move ventrally into the dorsal midline between E9.0–9.5, before moving laterally to the genital ridges during the E10–10.5 period (Molyneaux et al., 2001).

To successfully migrate to the genital ridge, PGCs need to survive the process and be highly motile in the hindgut (Molyneaux et al., 2001). Steel factor (SF) was shown to play a critical role in both survival and migration (Runyan et al., 2006). Very few PGCs that are in the gut at E10.5 are able to reach the genital ridge (Molyneaux et al., 2001), as downregulation of SF in the dorsal midline between E9.5 and E10.5 leads to apoptosis of PGCs that have failed to move laterally toward the genital ridges (Runyan et al., 2006). SDF-1/CXCR4 signalling and Wnt5a signalling is also crucial for the survival of germ cell and the directionality of migration (Molyneaux et al., 2003, Chawengsaksophak et al., 2012).

Around the time of reaching genital ridges (E10.5), mPGCs start proliferation again and initiate expression of several germline-specific markers such as Mvh (VASA)(Tanaka et al., 2000), Dazl (Lin and Page, 2005) and Sycp3 (Di Carlo et al., 2002). Those germline-specific genes are suggested to be methylation-sensitive (Maatouk et al., 2006), as their expressions is robustly silenced by methylation at CpG-dense promoters during post-implantation development but are activated during the global DNA erasure (Guibert et al., 2012). Deficiency of key germline-specific genes such as DAZL leads to defects at later developmental stages, such as meiosis, and overall failure in gametogenesis, especially spermatogenesis (Lin and Page, 2005, Saunders et al., 2003). After colonizing the developing gonad, female PGCs respond to retinoic acid (RA) signalling and initiate meiosis whereas male PGCs are responsive to FGF9 from newly specified Sertoli cells and are maintained by Nodal/Cripto signalling and Nanos2. In the absence of RA and presence of Activin/Nodal signalling, male PGCs enter mitotic arrest instead of meiosis (Sakashita et al., 2015, Spiller and Bowles, 2015) (Figure 1-1).

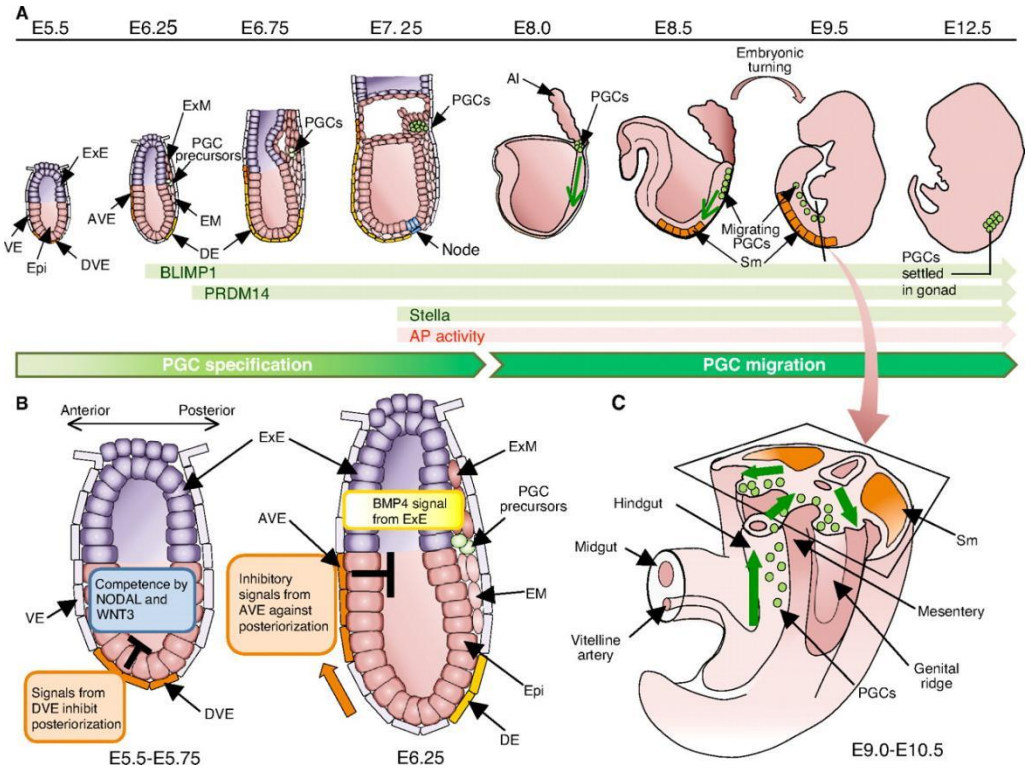


Figure 1-1 Development of mPGC (Saitou et al., 2012). (A) Illustration showing the overall development of mPGCs. A small number of mPGC precursors were detected at E6.25, expressing Blimp1 followed by Prdm14. The PGC population are determined and expands into ~40 cells expressing Stella at E7.25. PGCs start migration through hindgut from E7.5-E8 and settled in gonad at E12.5. (B) Multiple signals contribute to the specification and restriction of germ line lineage. NODAL and WNT signalling from VE are essential for the germline competence. BMP4 signal from ExE induce the PGC specification. (C) PGCs migrate towards the developing genital ridges through definitive hindgut endoderm. In hindgut, PGC move ventrally into the dorsal midline and then laterally to the genital ridges.

1.1.1.2 Key specifier and Signalling pathway for PGC specification

In mice, germ cell fate in the post-implantation epiblast is directly induced by Bmp4 originating primarily from the extraembryonic ectoderm (ExE) from E5.5, and by BMP2 from the visceral endoderm (VE)(de Sousa Lopes et al., 2007). WNT signalling is essential for the competency of germline lineage but cannot induce PGC specification without BMP signalling. Wnt3 expressed from around E5.75 at the posterior visceral endoderm (VE) and epiblasts

ensure the responsiveness of epiblast to Bmp4 (Ohinata et al., 2009). Through beta–CATENIN, WNT3 induces many mesoderm-associated transcription factors which may be activated before BMP4 in those epiblast-like cells, such as T (BRACHYURY) (Nikolic et al., Irie et al., 2015). This process is antagonized at the anterior side by several factors from anterior visceral endoderm (AVE), and Bmp8b from the ExE contributes to Bmp signalling by restricting AVE development. It has been shown that T (Brachyury) binds distinct regulatory elements of both Blimp1 and Prdm14 and directly upregulates these genes. However, without high-dose of BMP4, T will be prevented from activating Blimp1 and Prdm14, potentially by some WNT-induced inhibitors (Aramaki et al., 2013). BMPs induce SMAD1 and SMAD5 phosphorylation, dimerization with SMAD4 and translocation to the nucleus, which in turn induce the key transcriptional regulators of mPGCs, such as Blimp1 and Prdm14 (Saitou and Yamaji, 2012, De Felici, 2013, Sugawa et al., 2015). WNT signalling may also contribute to PGC specification by acting post-transcriptionally on those BMP signalling components. For example, by inhibiting GSK3-mediated phosphorylation of its linker region, WNT signalling can prevent Smad1 from degradation (Saitou and Yamaji, 2012). A recent study further suggests another important function of BMP signalling, which is to suppress *Otx2*, a roadblock for germline lineage (Zhang et al., 2018a). Nodal pathway is also critical for PGC specification, which is suggested to contribute to the optimal levels of Nanog during pre-PGC development and also to spatially confine the PGC niche (Senft et al., 2019). Neither Nodal null ES cells (Mulas et al., 2017) nor Smad2/3 double mutant ES cells (Senft et al., 2018) give rise to mouse PGC-like cells (mPGCLCs) in vitro. Eomes, which acts downstream of Nodal within the epiblast, is essential for mPGC development in vivo. mPGCLCs can be induced from Eomes null ES cells, albeit at lower efficiency. However, deletion of Eomes in the epiblast disrupts mPGC development. Although Blimp1 is still activated in pre-PGCs, those pre-PGCs fail to activate Ap2y, Nanog, Sox2 and Stella, suggesting that Eomes may not be required intrinsically for mPGC specification, instead, functions downstream of Nodal signalling to create a proper signalling environment for PGC niche formation (Senft et al., 2019).

Blimp1, induced by BMP4 at around E6.25, is generally a transcriptional repressor which is suggested to repress the somatic mesodermal differentiation program in approximately six cells in pre-streak epiblast cells (Ohinata et al., 2005). Blimp1-positive cells are considered lineage-restricted, i.e. PGC precursors. However, those six cells are unlikely to be the ancestors of the entire germ-cell lineage, as the increase of Blimp1-positive cells (i.e ~6 at E6.25 to ~40 Stella-positive cells at the base of the allantois at E7.25 (Lawson and Hage, 1994)) does not appear to be exponential. Therefore, the increase in PGC founder population is unlikely to be solely due to successive cell division (McLaren and Lawson, 2005). Additionally, lineage tracing experiments have shown that proximal epiblast cells, when exposed to BMP signals at E6.5, can also derive PGCs (Lawson and Hage, 1994, Tam and Zhou, 1996), suggesting the possibility of proximal epiblast cells contribute to the PGC population after E6.25 (McLaren and Lawson, 2005). On the other hand, Otx2 restricts mPGC population and safeguards somatic lineages. In Otx2-knockout embryos, mPGC exhibit increased number, and initial induction of mPGCLCs from Otx2-knockout pluripotent cells does not require cytokines such as BMP4 or Blimp1 activation (Zhang et al., 2018a). Shortly after *Blimp1, Prdm14* is expressed, and Tcfap2c (which encodes AP2γ) can be directly induced by BLIMP1(Murakami et al., 2016). Re-expression of multiple pluripotency genes, such as *Nanog, Oct4 (Pou5f1) and Sox2,* and epigenetic reprogramming are also induced in those putative germ cells.

It has been proposed that Blimp1, Prdm14 and AP2γ form a tripartite genetic network for PGC specification. Co-bindings or individuals of those factors may mediate upregulation of pluripotency and germ cell genes, repression of somatic programme and epigenetic reprogramming, contributing to PGC specification (Nakaki et al., 2013, Shirane et al., 2016, Yamaji et al., 2008). Also shown in an *in vitro* model, co-expression of BLIMP1, AP2γ and PRDM14 can sufficiently induce a PGC-like fate without cytokines, supporting the proposed tripartite genetic network (Magnúsdóttir et al., 2013, Nakaki et al., 2013). Shown by genome-wide DNA binding map, those key specifiers have in general different binding activities: BLIMP1 predominantly binds to promoters, whereas PRDM14 is enriched in distal regulatory elements (Magnúsdóttir et al., 2013). Targets of both BLIMP1 and PRDM14 can be

bound by AP2γ for either activation or repression (Tang et al., 2016). BLIMP1 and PRDM14 also show interaction with epigenetic modifiers such as SUZ12, histone deacetylases (HDACs), EHMT1 and protein arginine N-methyltransferase 5 (PRMT5), suggesting their potential role in initiating epigenetic reprogramming. The recruitment of epigenetic modifiers may also be required for modulating transcriptional activities (Saitou and Yamaji, 2012). For example, the transcriptional repression of somatic genes by BLIMP1 is at least partially dependent on recruitment of epigenetic modifiers such as Hdac3 and Setdb1, which removes active histone markers and catalyses repressive histone markers, respectively (Mochizuki et al., 2018b, Mochizuki et al., 2018a). The tripartite network is also suggested to have inhibitory effect on proliferation as early as E7.0 (Seki et al., 2007).

Expression of pluripotent genes in PGCs is also suggested to be essential for PGC specification and development. A recent study indicates that Nanog alone can induce mPGCLCs *in vitro* and might act cooperatively with BMP signalling *in vivo*, whereas Sox2 can inhibit PGCLC induction by Nanog (Murakami et al., 2016). It has been reported that deficiency of Nanog (Yamaguchi et al., 2009) or Oct4 (Kehler et al., 2004) in migrating PGCs leads to apoptosis. It has also been proposed that LIN28 which blocks the maturation of functional micro-RNA (miRNA) let-7 contribute to the activation of Blimp1 (West et al., 2009).

1.1.1.3 Resetting of epigenome during PGC development

Epigenetic reprogramming is a key event in PGC development**,** during which DNA methylation is erased and histone modifications are extensively remodelled. This resetting of the germline epigenome might play an essential role in PGC development as well as in reprogramming the developmental potential of PGCs to a totipotent state (Hackett et al., 2012a, Strome and Updike, 2015, Hill et al., 2018). Abnormal epigenetic states have also been linked to the formation of germ cell tumours (Killian et al., 2016, Sellers et al., 2018) and disruption in development (Hargan-Calvopina et al., 2016, Hill et al., 2018). Epigenetic reprogramming in the mammalian germline leaves little chance for any potential deleterious parental epigenetic information to be

transmitted, which partially explains the lack of evidence for transgenerational epigenetic inheritance in mammals (Heard and Martienssen, 2014). However, multiple studies in mammals reported inherited phenotypic variations, which are induced by environmental factors and cannot be explained by genetic variation (Carone et al., 2010, Braunschweig et al., 2012, Feeney et al., 2014, Chen et al., 2016c, Huypens et al., 2016, Bohacek et al., 2016). A recent study in hPGC further identified unique and repetitive regions that are resistant to global DNA demethylation, some of which are associated with metabolic and neurological disorders. Those 'escapee' regions are suggested to be potential carriers of parental epigenetic information and thus contribute to epigenetic inheritance (Tang et al., 2015). In view of this latest evidence, understanding PGCs biology can be an informative for understanding the basis of epigenetic inheritance in mammals (Daxinger and Whitelaw, 2012, Strome and Updike, 2015).

1.1.1.4 Distinct requirement of metabolism in PGC

Interplay between metabolism and epigenome has been discussed in stem cells (Harvey et al., 2019) and cancer (Miranda-Gonçalves et al., 2018). Various metabolites, such as acetyl-CoA, S-adenosylmethionine (SAM), NAD+ and α-ketoglutarate (αKG), are important donors for epigenetic modifications and cofactors for epigenetic regulators, and therefore cellular levels of metabolites may directly affect activities of key epigenetic regulators. In turn, epigenetic changes occurring on metabolism-related genes may affect their expression and consequently the metabolome.

It is not surprising that aside from the extensive reprogramming of epigenome, major changes in mitochondrial activity and metabolism have also been observed during PGC development. A recent paper investigated the metabolism in E13.5 mPGCs with both metabolomic information and transcriptional information, revealing high oxidative phosphorylation (OxPhos) activity and low glycolytic activity in gonadal mPGCs (Hayashi et al., 2017). Similarly, this metabolic shift from glycolysis to OxPhos was also observed in gonadal hPGCs, which was suggested to provide selective effects for new mutation of mitochondrial DNA (mtDNA) during PGC development (Floros et

al., 2018). On the other hand, oxidative to glycolytic switch occurs during the transition from naïve to primed pluripotency and therefore acquisition of competence for mPGC fate in vitro. Addition of αKG, a central metabolite of TCA cycle, is found sustaining PGC specification competence, potentially via preserving a specific epigenetic state (Tischler et al., 2019).

Apart from energy metabolism, vitamin C is also found to affect epigenetic reprogramming (Lee Chong et al., 2019). During the development of foetal germ cells, vitamin C is suggested to promote active DNA demethylation via enhancing TET activity. Deficiency of vitamin C in a mouse model leads to incomplete demethylation of meiosis regulators and transposable elements (TEs), which in turn results in reduced numbers of germ cells and delayed meiosis, resembling the phenotypes of *Tet1*-mutant embryos (DiTroia et al., 2019). Vitamin C can also enhance activity of other epigenetic regulators, including Jumonji-C domain-containing histone demethylases (JHDMs). Addition of L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (AA2P) was found to reprogram histone and also RNA methylation during oocyte maturation in pig (Yu et al., 2018b). A recent study further showed that addition of vitamin C enhanced efficiency of hPGCLC induction, along with increased expression of key PGC genes including *SOX17* and *BLIMP1* and increased levels of TET and 5-hydroxymethylcytosine (5hmC), the oxidized product of TET-mediated active demethylation (Li et al., 2019). This indicates that vitamin C, like energy metabolism, may affect PGC development from the very beginning.

Overall, the synchrony between metabolism and epigenome appear to be critical for proper development of PGC. A particular metabolome may be required for creating permissive epigenetic states for PGC specification and for promoting epigenetic reprogramming. The interplay between metabolism and epigenome may also provide explanations for intergenerational environmental effects on fecundity (DiTroia et al., 2019).

1.1.1.5 DNA repair mechanism and survival of PGC

Various genes are essential to PGC survival (Hamer and de Rooij, 2018). Here I focus on DNA repair components that have been investigated during PGC

development. DNA repair in germline is critical to ensuring that the fidelity of genome is maintained across generations. Failure in DNA repair, i.e. counteracting mutations, can lead to reduced number or even elimination of germ cells in offspring. Furthermore, when DNA damages in cell is beyond repair, failure in inducing apoptosis can also lead to devastating consequences such as neoplastic transformation of germ cells (Bloom et al., 2019).

Fanconi anaemia (FA) pathway, which is known for its role in repairing DNA interstrand cross-links (ICLs) lesions but also has other functions (Ceccaldi et al., 2016, Tsui and Crismani, 2019), is probably the most studied DNA repair mechanism in germline, as genetic mutations in several members of this pathway have been reported to lead to germline defects, including germline depletion (Kato et al., 2015). However, FA mutants, apart from Fancb and Fancl (Pog) mutants which show reduced mPGC numbers at E9.5, only show defects in gonadal mPGCs. The reduced PGC number in FA mutants involves ATM/TP53/p21 signalling as part of the DNA damage response and is considered to be due to proliferation defects rather than apoptosis (Nadler and Braun, 2000, Agoulnik et al., 2002, Luo et al., 2014). Interestingly, a more recent study shows that defects in FA pathway, i.e. Fanca and Ercc1 mutants, leads to accumulation of double-strand breaks (DSBs) and increased frequency of apoptosis in mPGCs from E10.5 (Hill and Crossan, 2019). Beyond FA mutants, mutations in a few more genes (Pin1 (Atchison et al., 2003), Mcm9 (Yunhai and C., 2015), Rev7 (Watanabe et al., 2013), Helq (Adelman et al., 2013)) have been correlated with PGC depletion. Most of them are also linked to FA pathway and/or repair of DNA ICLs. None of the mutants mentioned above affect mPGC specification. Also, most of them show no effect on the epigenome of mPGC, except Rev7(Mad2l2) and Fancb. It is suggested that Rev7 contributes to the switch from di-methylation of lysine 9 on histone H3 (H3K9me2) to tri-methylation of lysine 27 on histone H3 (H3K27me3), as Rev7-deficient mPGCs failed to downregulate H3K9me2 or upregulate H3K27me3 from E9.0 onwards (Pirouz et al., 2013). On the other hand, Fancb regulates accumulation of H3K9me2 positively but H3K9me3 negatively on sex chromosomes during meiosis (Kato et al., 2015).

Another DNA repair mechanism, base excision repair (BER) pathway, is also studied in germline as part of active DNA demethylation. A few components

of BER pathway show high expression in mPGC, cooccurring with active demethylation and BER pathway is suggested as one of the possible mechanisms to erasure Tet-oxidized products, i.e. 5hmC, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), in PGC (Hajkova et al., 2010, Wu and Zhang, 2017). DNA glycosylases that recognize and remove specific types of DNA bases, often damaged or mismatched, are considered the first step in standard BER pathway (Wood et al., 2005)(Figure 1-2). Among the 12 known mammalian DNA glycosylases, TDG was suggested as the most promising one for targeting Tet-mediated oxidative products (He et al., 2011). Many of those DNA glycosylases have shown binding activities towards 5-methylcytosine on DNA (5mC) and TET-oxidized products, including Mbd4, Ung, Neil1 and Neil3 (Spruijt et al., 2013). However, only TDG shows direct catalytic activities against 5fC and 5caC in vitro (Maiti and Drohat, 2011, Weber et al., 2016) (Figure 1-8). The function of TDG in active DNA demethylation has also been link to other DNA repair components including AID (Cortellino et al., 2011) and Gadd5a (Li et al., 2015). However, the function of TDG in DNA demethylation in vivo is hard to determine, as removal of 5fC in mouse pre-implantation embryos (Guo et al., 2014a, Shen et al., 2014) or PGCs (Hajkova et al., 2010, Yamaguchi et al., 2013a, Tang et al., 2015) appear to be independent with TDG. Other aspects of BER pathways that functions downstream of TET, such as whether a single or more nucleotides are replaced (i.e. short or long patch BER), remain unexplored.

Overall, the nature of DNA lesions and how they are responded with in PGCs remain largely unknown. PGCs appear to be very sensitive to DNA lesions, given that mutation in single component of FA pathway often leads to PGC depletion. It is tempting to speculate that DNA repair mechanisms, such as FA pathway and BER pathway, cooperate with epigenetic mechanisms to ensure proper epigenetic reprogramming and maintenance of genome stability during PGC development (Bokhari and Sharma, 2019, Kafer et al., 2016).

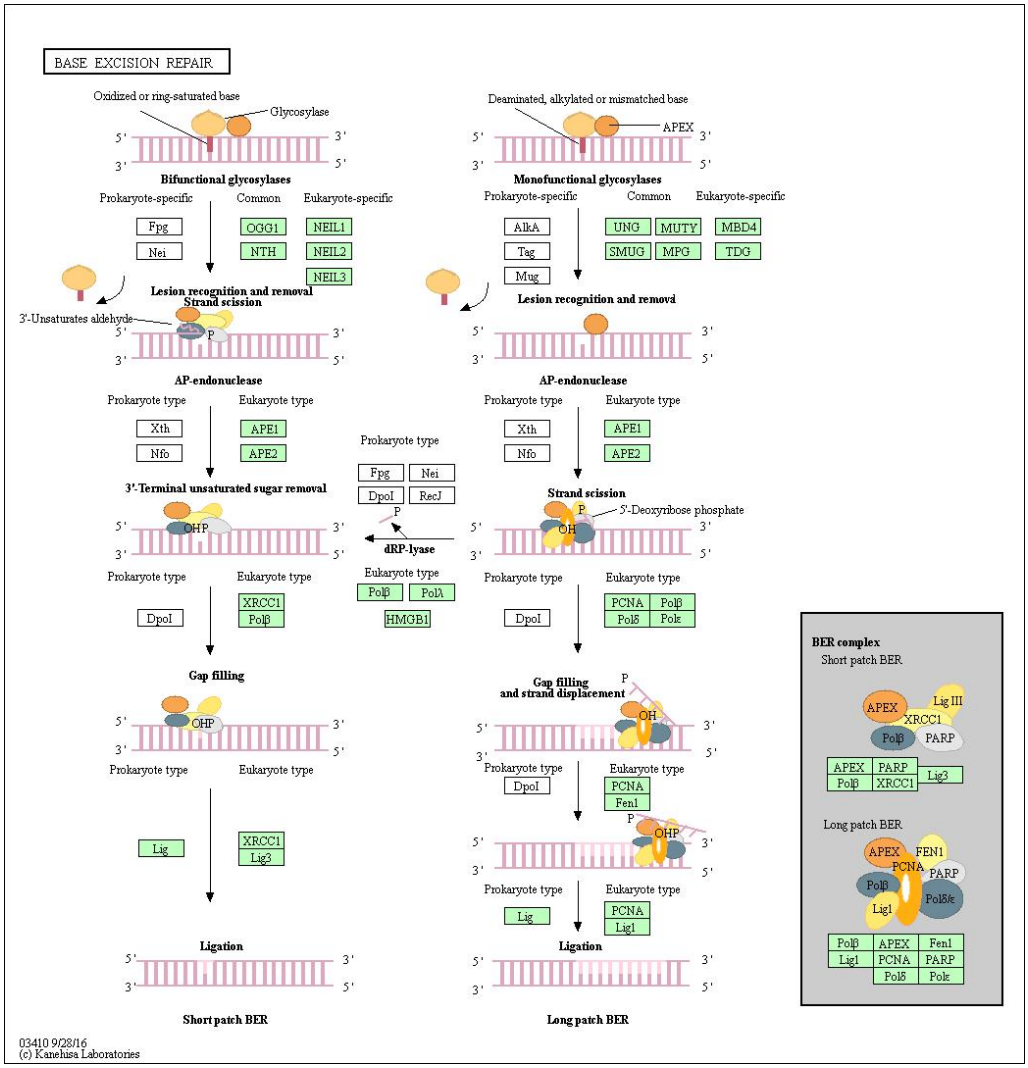


Figure 1-2 BER pathway mechanisms (Kanehisa and Goto, 2000, Kanehisa et al., 2018) showing the first step as glycosylase-mediated excision of DNA bases and differences between short and long patch BER.

1.1.2 Key difference between rodent and non-rodent PGC development

1.1.2.1 Specification of human and other non-rodent PGC requires different regulatory network compared to mPGC

Human embryo is not accessible around the timing of PGC specification (Week 2-3). Therefore, the origin, exact timing and mechanism of hPGC early development remain largely unknown. mPGCs are suggested to originate from

pre-gastrulation epiblast based on in vivo data (Saitou et al., 2002) and recapitulated by *in vitro* reconstitution (Hayashi et al., 2011). However, human PGC-like cells (hPGCLCs) are more efficiently induced from cells with early mesodermal character, suggesting it is possible that hPGC originate from the mesodermal precursors at the onset of gastrulation (Sasaki et al., 2015, Irie et al., 2015, Sugawa et al., 2015). Studies of PGCs of *Cynomolgus* Monkey (cyPGCs) in vivo and in 3D cultured embryos also provide another scenario that primate PGCs are derived from nascent amnion before gastrulation (Sasaki et al., 2016), as potential cyPGCs ($SOX17^+/TFAP2C^+$ or $SOX17^+/CDX2^+$) were detected within amniotic epithelium cells beneath the gastrulating epiblast 11-13 days post-fertilization (d.p.f) (Ma et al., 2019, Niu et al., 2019). This is further supported in hPGCLCs as the progenitors of hPGCLCs (i.e. D1 cells on the trajectory towards PGC lineage) express highly level of *TFAP2A* and show amniotic features). However, noting that those progenitors also show more consistent and evident gastrulating features (Chen et al., 2019b). Also, trajectory analysis revealed that cyPGCs from 3D cultured embryos show more similarity at transcriptional level towards gastrulating cells (Ma et al., 2019). Similarly, clustering analysis of scRNA-seq data revealed that early hPGCs from a rare gastrulating Carnegie Stage 7 human embryo (CS7 hPGCs) may emerge from the primitive streak (Tyser et al., 2020). Therefore, further investigations, with in vitro culture systems and in vivo embryos of pig and non-human primates, are warranted to characterise the $SOX17^+$ cells and the surrounding nascent amnion at molecular level, followed by lineage tracing of the fate of those cells.

Unlike the cylinder-shape mouse embryo, human embryonic disc does not have ExE, which is the main source of BMP4 signal for PGC specification in mice. Still, BMP signalling and WNT signalling are suggested to be the conserved signals for PGC specification in non-rodent mammals, based on studies in rabbit (Hopf et al., 2011), pig (Valdez Magaña et al., 2014, Kobayashi et al., 2017), *Cynomolgus* Monkey (Cy monkey) and human *in vitro* PGC systems (Sasaki et al., 2015, Irie et al., 2015, Kobayashi et al., 2017). Study in Cy monkey embryo suggests for primate PGC specification, WNT signalling (WNT3A) derives from the cytotrophoblasts and BMP signalling

derives from the amnion (BMP4) or extra-embryonic mesenchyme (BMP2) (Sasaki et al., 2016)

Robust *in vitro* inductions of hPGCLCs were first reported in 2015 (Sasaki et al., 2015, Irie et al., 2015, Sugawa et al., 2015), which are extremely useful for studying mechanisms of early hPGC development. Despite an overall similarity, induction of hPGCLCs has revealed several key differences between human and mPGC development. First of all, the core germline specification circuitry is different between rodent and non-rodent species. SOX17, a critical transcription factor for endoderm lineages which plays no role in mPGC specification, is distinctly expressed in those putative PGCs and was shown to be the key specifier of hPGC (hPGC)-like fate (Irie et al., 2015). SOX17 is also evidently detected in early cyPGC (Sasaki et al., 2016) and pPGC (Kobayashi et al., 2017). PRDM14 was demonstrated to have a different expression profile between mice and human. Although it is highly expressed in mPGCs and PGCLCs, in hPGCLCs it was reported to be strongly downregulated and maintain low (Sugawa et al., 2015). However, a more recent study detects clear PRDM14 signal in hPGCLCs and gonadal hPGCs and suggests an essential role of PRDM14 during hPGC specification. The role of PRDM14 in hPGC development is still divergent compared to in mPGC, as ChIP-seq analysis suggests PRDM14-targeted genes in hPGCLCs are largely different compared to mPGCLCs, and PRDM14 expression alone cannot induce germline lineage in human (Sybirna et al., 2020). pPGC also shows a different profile of PRDM14, as PRDM14 signal is weak and cytoplasmic in pre-migratory pPGC but clearly detected as nuclear signal in some early gonadal pPGC (Kobayashi et al., 2017). Another major difference is the absence of SOX2, the pluripotent gene critical for mPGC specification (Campolo et al., 2013), is absent in hPGCs and hPGCLC (Sasaki et al., 2015, Irie et al., 2015). An interesting study investigated the SOX2 and SOX17 genome occupancy in the SOX17-, SOX2+ embryonal carcinomas (ECs) and SOX17+, SOX2- seminomas, respectively. Both seminomas and ECs are suggested to arise from lesions of arrested PGCs (Berney et al., 2016). At transcriptional level, seminomas resemble PGCs, while ECs show similarity to ESCs. In general, SOX17-bound genes in seminoma cells and SOX2-bound genes in EC cells show large overlap, which includes germ cell and

pluripotency genes such as PRDM14, NANOG and LIN28A (Jostes et al., 2019). These results support the possibility that SOX17 in hPGC may replace some functions of SOX2 during mPGC development. The function of Nanog may also be different between mouse and hPGC specification, as unlike in mouse PGCLC induction, NANOG alone cannot induce hPGC fate in competent cells (Kobayashi et al., 2017). Both of NANOG and SOX2 are PRDM14-related genes, which also support a differential regulation system of PRDM14-related genes. Core regulatory factor BLIMP1 and TFAP2C are strongly expressed in both mPGC and hPGC, however in mPGC, it is suggested BLIMP1 acts upstream of TFAP2C, while in hPGC induction system, the initiation of TFAP2C is independent and upstream of BLIMP1 (Kojima et al., 2017).

Furthermore, germline competency is also differently regulated in human and mouse. Unlike in mice, knockout of T does not affect hPGCLC specification (Kojima et al., 2017). One the other hand, although Eomes is dispensable for mPGCLC induction, it is transiently expressed upstream of SOX17and is essential for hPGCLC induction and functions (Yokobayashi et al., 2017, Chen et al., 2017b, Kojima et al., 2017). SOX17-KO human induced pluripotent stem cells (hiPSCs) and incipient mesoderm-like cell (iMeLC) show similar level of EOMES with wild-type controls, while EOMES-KO iMeLC show reduced SOX17. Upon hPGCLC induction, EOMES-KO iMeLCs fail to upregulate SOX17 as well as BLIMP1. TFAP2C expression is initiated in EOMES-KO iMeLCs but fail to be maintained (Kojima et al., 2017), suggesting the initiation of TFAP2C depends on a distinct pathway. EOMES is detected in primitive streak and gastrulating cells of E12 Cy Monkey embryo, but not in TFAP2C-positve cyPGC, suggesting EOMES may be required for germline competency in primates but also downregulates after primate PGC speciation (Sasaki et al., 2016). As TFAP2C is a retinoic acid (RA)-responsive gene (Hoei-Hansen et al., 2004), and EOMES is suggested to upregulate retinoic acid receptor (RAR) signalling pathway during hPGCLC induction (Kojima et al., 2017), an interesting hypothesis is proposed that after induced by BMPs, TFAP2C expression is first maintained by EOMES-mediated RA signalling pathway and later by feedback loops formed within core regulatory circuitry of hPGC development (Jostes and Schorle, 2018).

First described as a gatekeeper of PGC specification in mice, the role of OTX2 has not been investigated during hPGC specification (Zhang et al., 2018a). Expression of OTX2 is downregulated in WT hPGCLCs compared to human pluripotent stem cells, but not in Blimp1-mutant hPGCLCs (Tang et al., 2015, Kojima et al., 2017). However, OTX2 shows slightly higher expression level in gonadal hPGC compared somatic cells (Tang et al., 2015), and a higher expression of OTX2 is detected in testicular germ cell tumours (TGCT) tissues, which resemble PGCs and pluripotency cells, compared to normal testis tissue. OTX2 is also identified as a direct target of SOX17 in seminoma cells. Therefore, it remains largely unclear if and when the downregulation of OTX2 is essential for hPGC specification and development.

It is worth noting that PGCs of plains vizcachas (*Lagostomus maximus*) (lmPGC), which are basal rodents, express SOX17 before migration, together with OCT4 and germline markers such as STELLA and FRAGILIS. However, BLIMP1 was not detected in lmPGC at any examined stage, suggesting a largely divergent regulatory network of PGC specification compared to mouse. Early embryo of *l.maximus* also shows a flat, disc shape instead of an egg cylinder, and no contact between epiblast and the ExE is observed before or after gastrulation. It is possible that the specification of PGC in mouse and rats, which depends on signalling from ExE and lacks Sox17, is a newly evolved trait, for the superfamily *Muridae,* the most recently evolved group of rodents (Leopardo et al., 2018, Leopardo and Vitullo, 2017).

1.1.2.2 Different tempo and mechanisms of epigenetic reprogramming in hPGC

Despite the general similarity of epigenetic reprogramming in mouse and hPGC development, several evident differences have been reported. Global DNA demethylation is one featured event of epigenetic reprogramming, which is suggested to occur in two phase in mPGCs, with global loss of DNA methylation via replication-dependent passive way before the gonadal stage, and locus-specific loss of DNA methylation via Tet1/2-mediated active demethylation (Seisenberger et al., 2012). For example, parent-of-origin imprints in mPGCs are mainly erased via the active mechanism after mPGC

reach the gonads (Hackett et al., 2013, Kawasaki et al., 2014). However, recent studies in hPGCs *in vitro* and *in vivo* indicates 5mC level on imprints is largely reduced with enrichment of 5hmC before genital ridge colonization, suggesting a different 'tempo', and potential mechanism of DNA demethylation in human and mPGCs. Profiles of histone modifications during hPGC development (Tang et al., 2015), such as H3K27me3 and H3K9me2, also show difference compared to histone remodelling in mPGCs (Seki et al., 2007, Kagiwada et al., 2013, Seki et al., 2005). With different profiles of DNA methylation and histone modification, it is not surprising that the timing of X chromosome reactivation (Chuva de Sousa Lopes et al., 2008, Sugimoto and Abe, 2007, Guo et al., 2015) and chromatin accessibility also show difference between human and mouse germ cell development (Guo et al., 2016) (Figure 1-3).

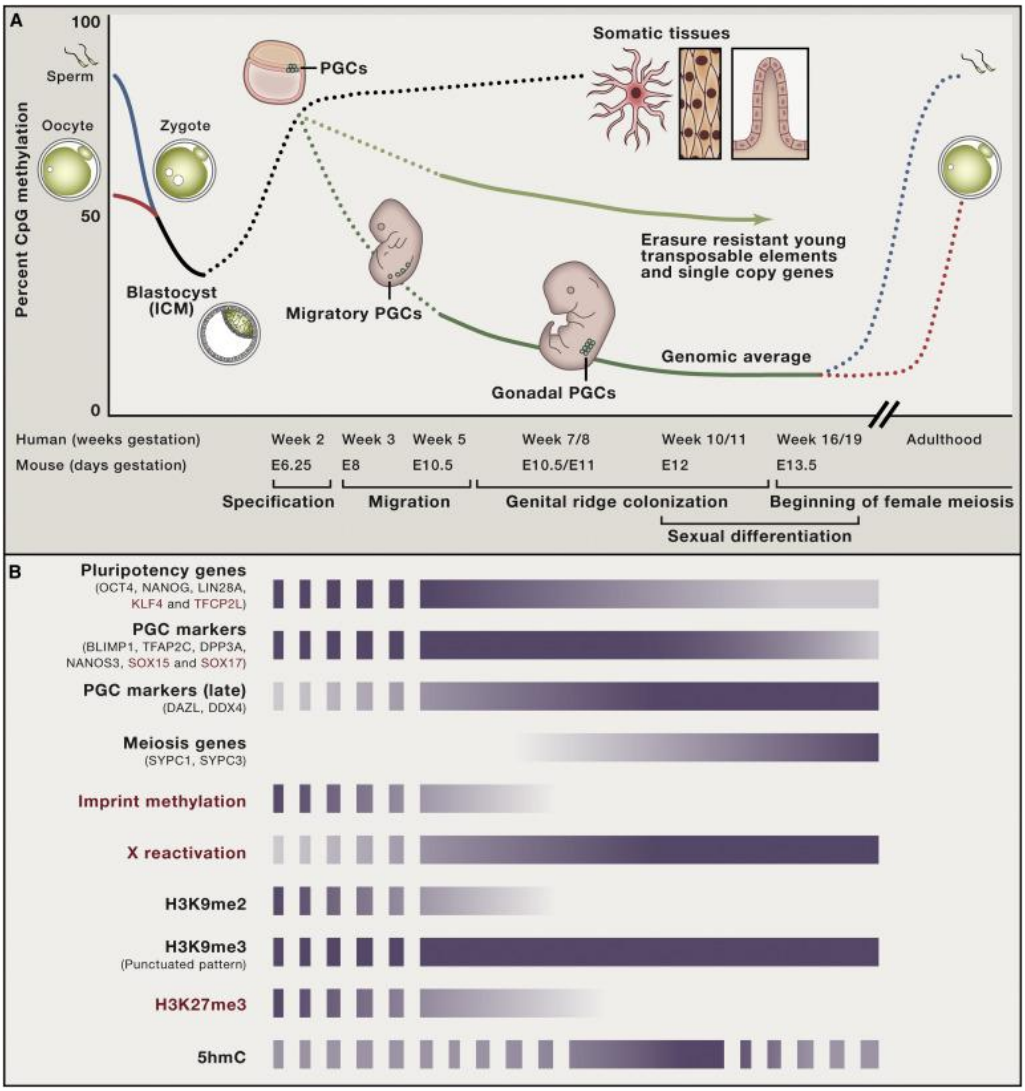


Figure 1-3 Epigenetic reprogramming in hPGC (von Meyenn and Reik, 2015).

(A) Two waves of global demethylation also occur in human lifecycle. One wave starts after fertilization where the epigenome from paternal (blue) and maternal (red) gametes are reset by the blastocyst stage and the other in hPGCs following the de novo methylation in the epiblast. In Wk7-9 hPGCs, DNA methylation reaches the basal level and significant loss of methylation has occurred genome wide. However, a small number of demethylation resistant loci, including some evolutionarily young TEs and single copy genes, retain partial methylation and are potential carriers of transgenerational epigenetic information. (B) Similarities and differences (marked by red) have been observed between human and mouse PGC development. hPGCs show high expression of naïve pluripotency genes KLF4 and TFCP2L and require SOX17 for specification. Imprint methylation and X reactivation in hPGCs are

evident at Wk4, which is suggested to be earlier compared to in mPGCs. Also, unlike in mPGCs, H3K27me3 later becomes absent in hPGC after the enrichment in migratory and early gonadal stages.

1.1.2.3 Pig PGC development shows similarity to hPGC

Unlike the cylinder-shaped embryo in mice, pig develops an embryonic disc. Development of post-hatching porcine embryo (E8-15) was summarized in Table 1-1 (Contreras Caro del Castillo, 2011). A spherical porcine embryo which is newly hatched from zona pellucida (E8-9) contains well-defined inner cell mass (ICM), extending hypoblast and the trophectoderm (TE). The ICM is covered by the polar TE (i.e., Rauber's layer). During the pre-streak stage I (E10-11) the hypoblast is fully formed (i.e., The hypoblast covers the inside of both ICM and trophectoderm) and the Rauber's layer degenerates. The remaining cells of the former ICM are now referred to as the epiblast and are exposed to the uterine cavity (Vejlsted et al., 2006a). At the following pre-streak stage II or early primitive streak (EPS) stage (E11-12), the conceptus initiates elongation and the spherical embryo becomes ovoid. A crescent-shaped thickening is observed within the caudal 1/3 of the porcine embryonic disk at this stage, which is due to the delamination of extraembryonic mesoderm (Fléchon et al., 2004). The primitive streak (PS) stage (~E12) is characterised by the occurrence of primitive streak with an anterior node. Cells around the former crescent-shaped thickening appear to gather in the midline to start their epithelial-mesenchymal transition (EMT). At this stage, amniotic folds are visible at the borders of the embryo proper. Neurulation also takes place from PS stage and leads to the formation of neural plate. The folding of neural plate via differential growth forms the neural groove, which is the feature of neural groove stage (E13-14). At the neural groove stage, the embryo proper is partly enclosed by the amniotic folds. The neural groove extends from the anterior towards the posterior pole of the embryo proper. When the neural groove almost reaches the posterior pole in porcine embryo,

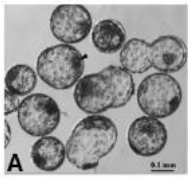
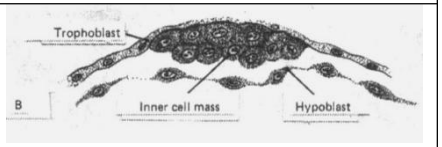
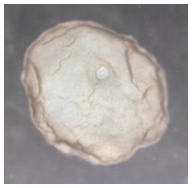
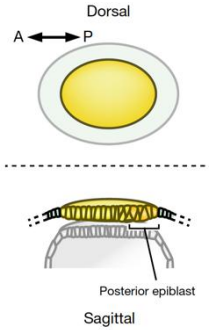
the paraxial mesoderm thickens to form somites which defines the somite stages (Vejlsted et al., 2006a, Vejlsted et al., 2006b, Oestrup et al., 2009).

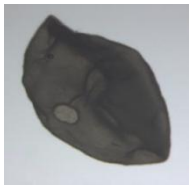
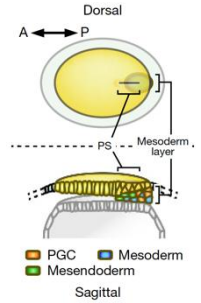
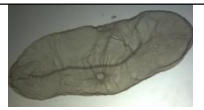
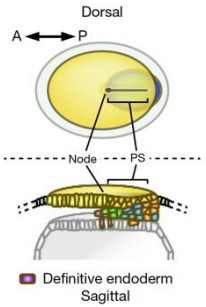
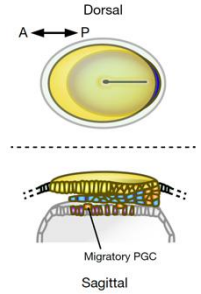
A recent study indicates that the posterior pre-streak epiblast, which shows expression of mesodermal factors T, is most competent for pig PGC (pPGC) induction. In response to WNT and BMP signalling, pPGC is induced and shows sequential upregulation of SOX17 and BLIMP1. Specified pPGCs (i.e., $SOX17^{+}$ and $BLIMP1^{+}$) were observed at the posterior end as early as the EPS stage (E11.5-12). The stage of competence and signalling requirement in pPGC specification is mostly recapitulated in human and monkey *in vitro* systems. At E14-14.5, a distinct pPGC cluster is observed at the posterior end (Figure 1-4). Interesting, it is suggested that the expansion of pPGC population is due to recruitment of competent cells, but not proliferation, based on the absence of de novo DNA in newly specified pPGCs and their precursors (Kobayashi et al., 2017). The cease of proliferation also progressively occurs in mPGCs but mostly after specification and become evident at early migration stage, and they are reported rapidly proliferating until E7.25 (Seki et al., 2007, Ohno et al., 2013). The way of expansion in hPGC is largely unknown. During the induction of hPGCLCs, downregulated genes are enriched in GO term 'positive regulation of cell proliferation' while upregulated genes enriched in 'negative regulation of cell proliferation' (Sasaki et al., 2015), yet a large proportion of hPGCLCs are positive for proliferating marker Ki-67 (Irie et al., 2015).

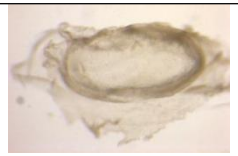
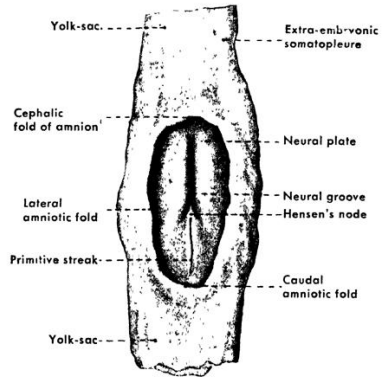
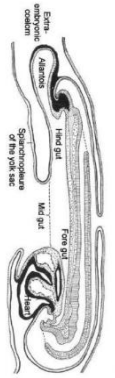
Migratory pPGCs can be observed as early as around E12.5-13.5 (Kobayashi et al., 2017). The migration pattern is largely shared between mouse and pig. A large number of migratory pPGCs appear in hindgut from 8-somite stage (Figure 1-5). However, the precise timing is complicated to compare. pPGCs appear less synchronous and more randomly distributed during the migration compared to in mice. The mechanism underlying the migration event in the pig embryo remains largely unknown (Hyldig et al., 2011b). At E22, pPGCs have reached the genital ridges which will further develop into primitive gonads (Figure 1-6) (Hyldig et al., 2011a).

The transcriptional profile of pPGC is largely similar to hPGCLC, with SOX17 as the key specifier and downregulation of SOX2. There is a difference in the re-expression of NANOG in the mouse, whereas it is constantly present in the posterior regions that give rise to pPGCs from E9.5 (Kobayashi et al., 2017). Epigenetic profiles of nascent pPGC are largely different to the ones in mPGCs, specifically the significant reduction of 5mC accompanied with early enrichment of 5hmC during the cease of proliferation at pre-migratory stage, suggesting a major role of active demethylation in early pPGC development (Figure 1-7). The erasure of imprints in pPGCs also appears to be earlier compared to mPGC (Hyldig et al., 2011a). Active demethylation observed in nascent pPGC could be an explanation for the active erasure of imprints prior to genital ridge colonization in hPGCs. Similar to hPGC *in vitro* and *in vivo*, H3K9me2 and its contributor G9a are downregulated in pig pre-migratory and early migratory PGCs. The other histone modification, H3K27me3, show enrichment in migratory and early gonadal pig PGC, followed by large downregulation in the following stages. A similar profile is also observed in hPGC *in vivo*. In contrast, high level of H3K27me3 is maintained in mouse gonadal PGCs (Hyldig et al., 2011a).

Table 1-1 This table is modified from Contreras Caro del Castillo, 2011. The images are modified from Vejlsted et al., 2006a and Contreras Caro del Castillo, 2011. The illustrations are modified from Patten, 1931 and Kobayashi et al., 2017. This table shows day of collection after fertilization, and morphological characteristics of embryonic and extra-embryonic tissues. Inner cell mass (ICM); Early primitive streak (EPS); Primitive streak (PS); Late primitive streak (LPS), Early neural groove (ENG), Neural groove (NG), Somite Stage (STS).

Stage	Days after AI	Conceptus Shape	Embryo Shape	Embryonic tissue	Extraembryonic tissue	Bright-field image	Illustration
Expanded hatched blastocyst	8-9	/	Spherical	ICM	Hypoblast/Primitive Endoderm Trophoblast		
Pre-streak stage I	10-11	Spherical	Round-disc	Epiblast	Hypoblast/Primitive Endoderm Trophoblast		

Pre-streak stage II/EPS	11-12	Ovoid	Oblong/ Thickening crescent in the posterior pole	Epiblast	Hypoblast/Primitive Endoderm Extraembryonic-mesoderm Trophoblast		
PS	12	Tubular	Primitive streak	Ectoderm	Hypoblast Primitive Endoderm Extraembryonic-mesoderm Trophoblast		
LPS	12-13	Tubular-Elongated-early filamentous	Oblong-Primitive streak	Ectoderm	Hypoblast Extraembryonic-mesoderm Trophoblast		

ENG	13	Elongated-Filamentous	Oblong-pear shape-Primitive streak	Ectoderm Mesoderm Endoderm	Trophoblast Amnion		
NG	14	Filamentous	Pear-shape elongated	Ectoderm Mesoderm Neural plate Neural groove	Trophoblast Amnion		
STS	From 15	Filamentous		Somites Anterior neuropore	Trophoblast Amnion Yolk sac Allantois		

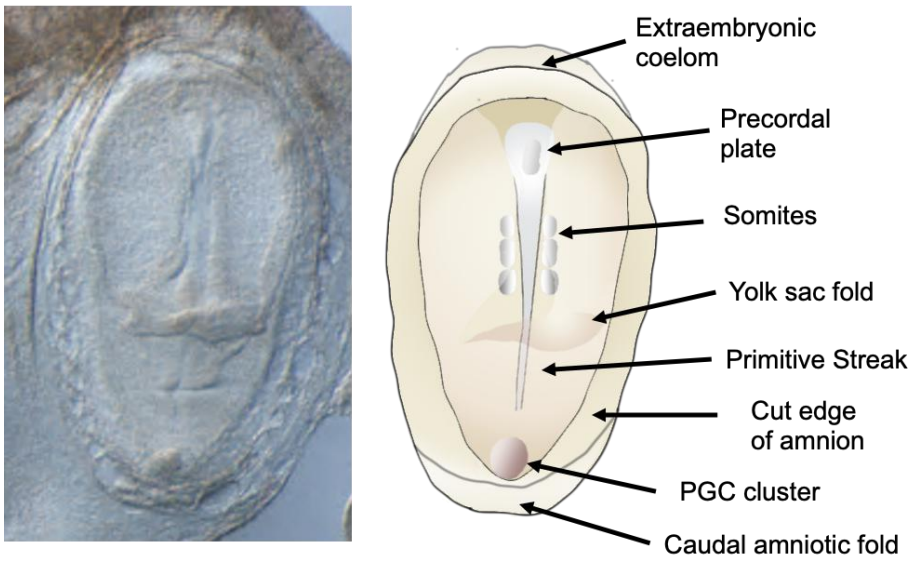


Figure 1-4 Bright field image (left) and illustration (right) of a 3-somite porcine embryo (~E14). An evident PGC cluster is present at the posterior end of the embryo.

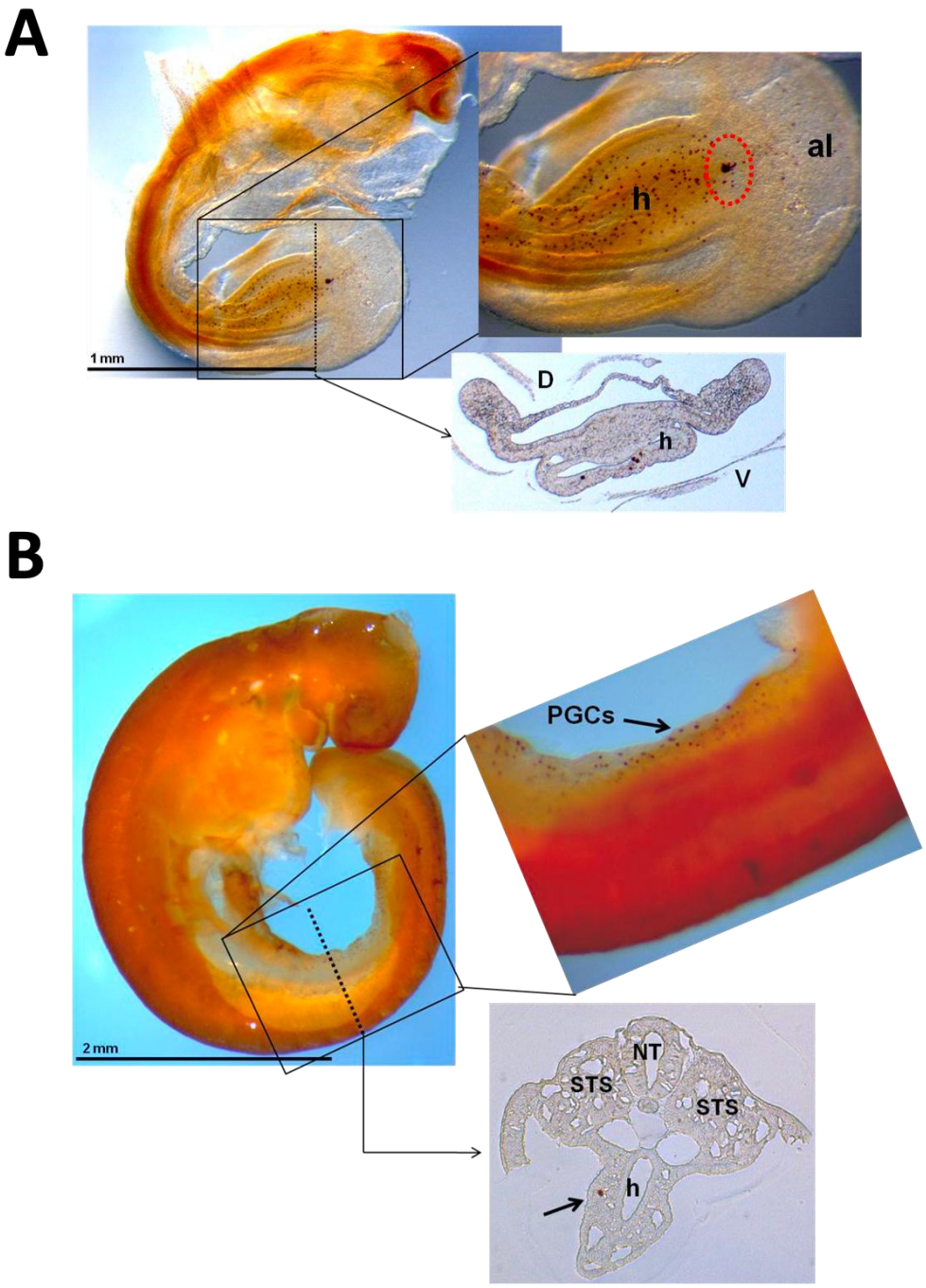


Figure 1-5 Whole-mount images and transversal sections of an 8-somite (A) and a E17 porcine embryo (B). pPGCs were stained with OCT4 antibody. Migratory pPGCs can be observed in the hindgut. Red circle in (A) indicates the pPGC cluster. Black arrow in (B) points at a single PGC in the hindgut. (h) hindgut, (al) allantois, (V) ventral, (D) dorsal. (NT) neural tube, (STS) somites. Modified from Contreras Caro del Castillo, 2011.

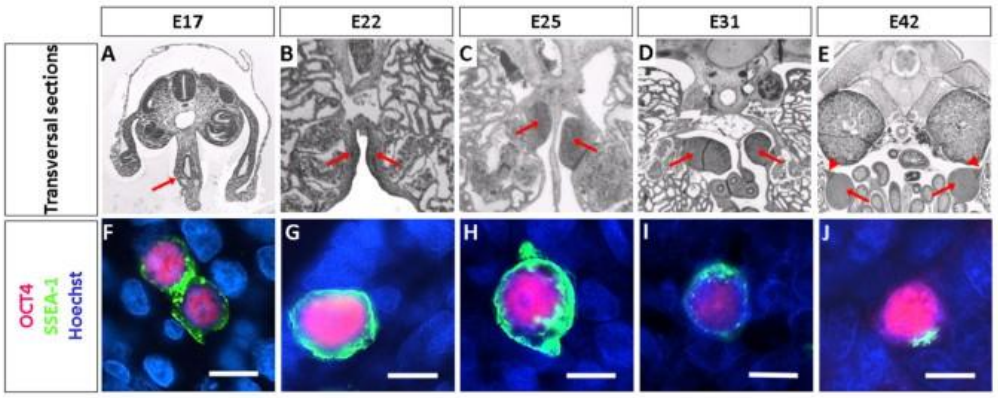


Figure 1-6 The top panel shows transversal sections of porcine embryos (E17-42). pPGC are found in hindgut of E17 embryo (A) and genital ridges or primitive gonads of E22-42 embryo (B-E). Arrows indicate the PGC containing tissue. Arrowhead depicts the mesonephric connection. The bottom panel shows double IF staining of the OCT4 and SSEA-1 in transversal sections of porcine embryos of the above stages (Hyldig et al., 2011a).

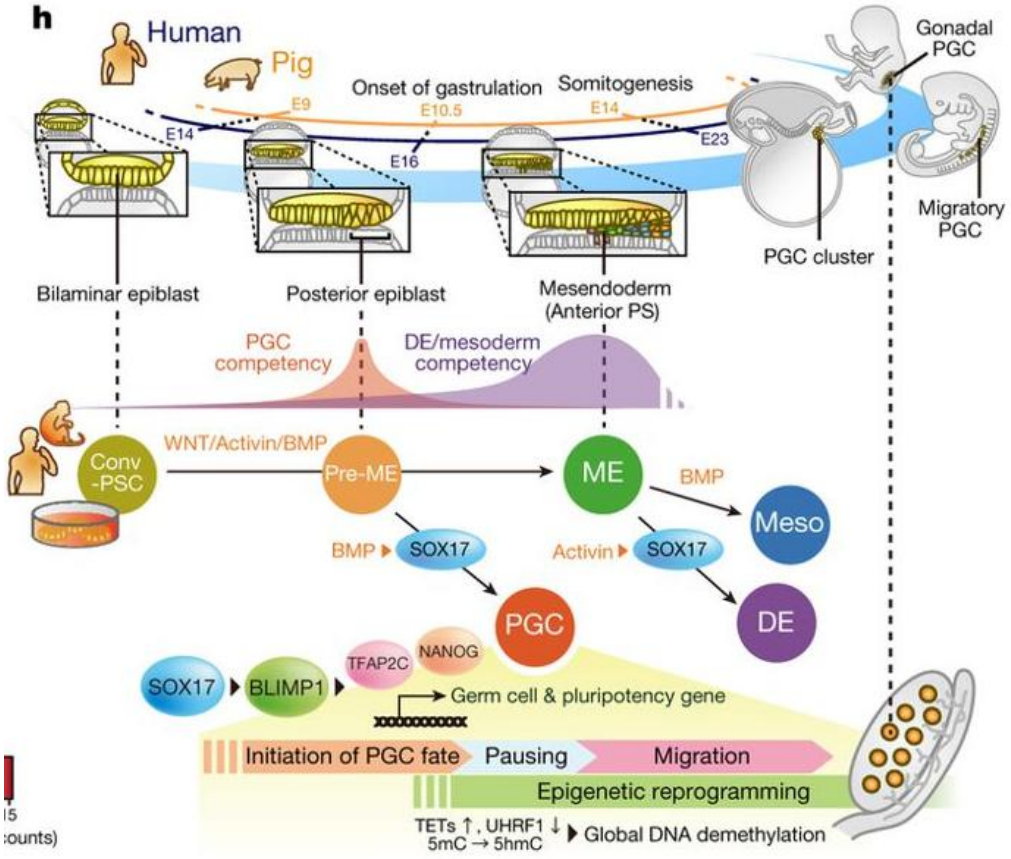


Figure 1-7 In vivo pPGC specification shows similarity to hPGC and cyPGC in vitro induction (Kobayashi et al., 2017). The epiblasts of three species develop in flat disc embryos, where SOX17 likely regulates both PGCs and DE. In response to BMP and WNT signalling, SOX17 and BLIMP1 are upregulated which are necessary and sufficient for inducing PGCs, and for initiating the germline-specific epigenetic program such as global DNA demethylation.

1.2 Epigenetic reprogramming in mammalian development

Currently, global DNA demethylation appears to be the most-studied process during epigenetic reprogramming in mammalian development. However, epigenetic modifications are highly cooperative, and systems which regulate DNA methylation, histone modification and ncRNAs are very likely to work in parallel to guarantee the robustness and high fidelity of epigenetic reprogramming while maintaining the cell identity and functionality of

individual cells. Understanding regulatory mechanisms underlying different epigenetic modifications and their cooperation during epigenetic reprogramming will lead to a more comprehensive picture of epigenetic reprogramming (Hackett et al., 2012b). Therefore, the potential functions of several key histone modifications in PGC development will also be discussed.

1.2.1 **DNA modification during mammalian development**

1.2.1.1 5mC and Demethylation

The presence of 5mC was first reported in mammals in the late 1940s and its role as an epigenetic marker has been widely studied and documented (Plongthongkum et al., 2014, Jones, 2012, Breiling and Lyko, 2015). Cytosine methylation mostly occurs in a CpG dinucleotide context, yet CpG islands (CGIs), which are characterized by a very high CpG density, are typically hypomethylated. CGIs are found in approximately 70% of mammalian gene promoters (Saxonov et al., 2006). Other gene regulatory elements such as enhancers may also contain CGI.

5mC on DNA can be catalysed by three conserved DNA methyltransferases (DNMT), DNMT1, DNMT3A and DNMT3B as well as one newly discovered and mouse specific Dnmt3c which specifically targets evolutionary young TEs in male germline (Barau et al., 2016). As Dnmt3c is currently only annotated in a few rodent species, it is difficult to predict whether it exist in non-rodent species and therefore this gene is not further discussed here. De novo methylation via Dnmt3a/b at appropriate genomic regions, such as imprinted genes, will be faithfully maintained by Dnmt1/Uhrf1 complex during DNA replication and cell division. Dnmt3a/b are also required for robust maintenance of methylation (Chen et al., 2003, Arand et al., 2012). De novo methylation by Dnmt3a/b most profoundly occurs in two phases in early development, following two waves of global demethylation in pre-implantation embryo and developing PGCs, respectively. The specificity of Dnmt3a/b to methylate the appropriate regions of the genome is a critical mechanism in establishing epigenome in both developing individual and their germline which contribute to next generation.

Although DNMT3A and DNMT3B (full-length isoform) are highly homologous de novo methyltransferases, they have been reported targeting distinct genomic regions. Still, mechanisms underlying the specific targeting of de novo methylation are largely unclear. It is unlikely that specificity of de novo methylation is solely dependent on recognition of specific genomic sequence by DNMT3A/B. Other factors, such as DNA binding proteins, interference of transcriptional machinery, the effects of histone modifications and nucleosome positioning, may also help shape the de novo established epigenome (Rose and Klose, 2014). Dnmt3L, which lacks a catalytic domain compared to Dnmt3a/b, is considered a cofactor for de novo methylation (Guenatri et al., 2013). Dnmt3L has been reported with positive (Ooi et al., 2007, Jurkowska et al., 2011a, Jurkowska et al., 2011b) and negative effect (Neri et al., 2013b) on DNA methylation under different context. Although Dnmt3L is suggested to be important for mPGCLC induction (Neri et al., 2013b) and involved in the de novo methylation in mouse germ cells (Smallwood et al., 2011, Bourc'his and Bestor, 2004) and early embryo (Guenatri et al., 2013), its expression is controversial in hPGCLCs (Sasaki et al., 2015, Irie et al., 2015, Sugawa et al., 2015) and barely detected in gonadal hPGCs (Tang et al., 2015, Gkountela et al., 2015, Guo et al., 2015) or early pig PGCs. It is not detected in the human oocytes (Okae et al., 2014) or the oocyte and early embryo of pig and cow (Ivanova et al., 2020). However, it's upregulated in early and gonadal cyPGCs (see section Chapter 22.32.3.22.3.2.1 Figure 2-35), of which the function warrants further investigation.

Dnmt1 is considered the 'maintenance' methyltransferase because of its higher affinity for hemi-methylated DNA (Jeltsch and Jurkowska, 2014). Uhrf1, a recruitment factor, is required for the specific targeting of DNMT1 to hemi-methylated DNA. The DNMT1/Uhrf1 complex recognises hemi-methylated CpG sites and modifies the unmethylated daughter strand with the same methylation patterns that originated in the parental strands.

During epigenetic reprogramming, the removal of 5mC can be achieved by inhibition of DNMTs and their partners, which allow dilution of 5mC through DNA replication (also known as passive demethylation), or by oxidation by TET proteins (TET1/2/3) into 5hmC, 5fC and 5caC (Figure 1-8). In early

embryos, classic models suggest the male pronucleus undergoes active demethylation mediated by TET3. 5hmc and other oxidative products generated by TET could be removed via base excision repair (BER) pathway as well as replication-dependent dilution (Gu et al., 2011, Inoue and Zhang, 2011). Slower demethylation kinetics is observed in maternal genome, as maternal genome is tightly bound by STELLA and therefore protected from TET3-mediated active demethylation (Nakamura et al., 2007, Messerschmidt, 2012, Nakamura et al., 2012), but still undergoes replication-dependent passive demethylation caused by nuclear exclusion of majority of DNMT1 (Hirasawa et al., 2008, Inoue and Zhang, 2011). After zygotic stage, TET3 expression is rapidly reduced at 2-cell stage (maternal-zygotic transition) and replaced by TET1/2 around morula stage (Gu et al., 2011, Cao et al., 2014, Lee et al., 2014a, Gao et al., 2017), while the genome continues to undergo demethylation until the blastocyst stage. DNA re-methylation is restored during embryonic lineage establishment (Inoue and Zhang, 2011). New single-base-resolution techniques for DNA methylation profile provide more detailed information in both mouse and human which partially disagrees with the classic model and indicates that Tet-mediated demethylation is also evident in the maternal genome (Wang et al., 2014, Shen et al., 2014, Tsukada et al., 2015, Guo et al., 2014a). A recent study further suggested that in the mouse paternal genome TET3 may function as an antagonist to de novo methylation activity in later zygotic stages, but not necessarily required for demethylation before PN3, as TET3-independent 5mC erasure occurs before evident 5hmC generation, and maternally inherited DNMT3A and DNMT1 can newly deposit 5mC on paternal genome around PN3 (Amouroux et al., 2016). Also, deletion of maternal Tet3 did not seem to affect the pre-implantation development(Gu et al., 2011), but do reduce rate of full-term development and neonatal development (Tsukada et al., 2015), which may relate to defects in paternal gene activation and activity of pluripotency genes NANOG (Gu et al., 2011, Costa et al., 2013, Ito et al., 2010). In mice, it is also suggested that partial impairment in the maintenance of symmetric DNA methylation at CpG dyads, rather than active demethylation, is the main driver of DNA demethylation in PGCs and in both paternal and maternal genome of the zygote (Arand et al., 2015). The link between increased cell numbers (i.e. enhanced cell replication)

and decreased DNA methylation has been reported in studies of assisted reproductive technologies. IVF and SCNT embryos have been often observed with higher global DNA methylation level, while treatments, such as addition of *Xenopus laevis* egg extract (Gillespie et al., 2012) or reproductive fluids (Canovas et al., 2017) in the media, that improve the blastocyst rate are often found to decrease DNA methylation and increase cell number per blastocyst.

Although the profiles in human and mice are generally similar, distinct features were observed between 2-cell to morula stage (Guo et al., 2014c, Smith et al., 2014). The profiling in zygotic stages of human embryos is obtained via immunocytochemistry, which suggest co-existence of 5mC and 5hmC in both pronuclei in a non-reciprocal pattern (Petrussa et al., 2016). This non-reciprocal pattern is also reported in human oocytes (Petrussa et al., 2016) and mouse zygotes (Salvaing et al., 2012), which is partially consistent with the finding of de novo methylation at PN3 (Amouroux et al., 2016). DNA methylation dynamics has also been studied in mammalian species other than mice and human. In rhesus monkey, an unexpected methylation peak at 8-cell stage and reduced represented bisulfite sequencing (RRBS) indicates, unlike the unidirectional demethylation observed in mice and human, de novo methylation is concurrent with demethylation during monkey early embryogenesis (Gao et al., 2017). DNA methylation and hydroxymethylation profiling of early embryos in other species, including bovine (Wossidlo et al., 2011), porcine (Cao et al., 2014, Lee et al., 2014a), ovine (Jafarpour et al., 2017, Masala et al., 2017) and equine(Heras et al., 2017), are mainly obtained via immunochemistry, which can separate 5mC and 5hmc but do not have single-base resolution. Profiles of 5mC, 5hmC and their modifiers seem to be species-specific and considerably affected by IVF and SCNT, which warrant further investigation into global demethylation of early embryos in multiple mammalian species (Cao et al., 2014, Zhang et al., 2016b, Canovas et al., 2017).

During the global demethylation in pre-implantation embryo, methylation at imprints, retrotransposons as well as a portion of differentially methylated CGIs in oocyte-specific and sperm-specific methylated regions are at least partially maintained (Kobayashi et al., 2012, Smith et al., 2012, Smith et al., 2014). Those maintained methylated regions could be potential carriers of

parental epigenetic information and contribute to epigenetic inheritance. Several factors have been shown to contribute to methylation maintenance at specific loci, including DNMT1 and DNMT1o (the oocyte-specific form of DNMT1), STELLA/PGC7 (Wossidlo et al., 2011), TRIM28/ZFP57 as well as other KRAB family members with similar properties with ZFP57 (Strogantsev et al., 2015, Quenneville et al., 2011, Kumar et al., 2017).

The maintenance or erasure of methylation at imprints is one of the key differences between global demethylation in embryos and in PGCs. Based on DNA methylation and expression profiles obtained from mPGCs, global demethylation in pre-gonadal PGCs is mediated by passive DNA demethylation caused by repression of DNMT3A/B and UHRF1, whereas after reaching genital ridge, TET1/2 erase methylation at specific regions such as imprinted regions. The function of TET1/2 on imprint erasure has been shown in several studies (Dawlaty et al., 2013, Yamaguchi et al., 2013a, Yamaguchi et al., 2013b, Hackett et al., 2013). TET1 may also be involved in regulation of meiosis, as although genome-wide demethylation in primordial germ cells is not largely affected, Tet1 deficiency leads to defective DNA demethylation and decreased expression of a subset of meiotic genes (Yamaguchi et al., 2012). However, fertility of TET1/2 single KO and double male KO mice are overtly normal, which lead to questions about the functional redundancy and specificity of TETs as well as the necessity of TET1/2 during PGC demethylation and development. Alternatively, TET1/2 and 5mC oxidation may be required to “kick-start” the demethylation process (Hahn et al., 2014).

It is worth noting that a new type of stem cell, named expanded potential stem cells (EPSC), has been cultured from individual mouse eight-cell blastomeres, with inhibitors targeting MAPKs, Src, Wnt/Hippo and Tnks1/2. EPSCs can also be converted from mESCs and iPSCs. Distinct from naïve mESCs, EPSCs can contribute to the embryo in chimeras, and differentiate towards trophectoderm and extraembryonic endoderm lineages. DNA modification 5mC, and 5hmC, as well as histone modification H3K4me3, H3K27me3 and H3K27ac, has been characterized for EPSCs. Although scRNA-seq indicates the transcriptome of EPSCs may be closer to the ones of other in vitro culture pluripotent stem cell rather than of in vivo eight-cell blastomere, EPSCs may provide novel material for studying epigenome, especially histone

modification and transcriptional factor binding, in pre-implantation embryos (Yang et al., 2017a).

The overall DNA demethylation dynamics are similar between mouse and human foetal germline. DNA methylation is recognized as a principal contributor to the stability and regulation of gene expression for differentiation and maintenance of cellular identity. The methylome of gonadal PGCs reaches its lowest level at E13.5 in mouse and Wk9 in human, with most promoters and retrotransposons demethylated, yet gene expression is under control and genome integrity is protected, indicating the global uncoupling of DNA methylation and transcription during epigenetic reprogramming in PGCs. However, in hPGCs, demethylated promoters of some genes related to meiosis, KRAB-ZFP and PIWI pathway, show correlation with increased expression. Similar correlation is also observed in early gonadal mPGCs (Hargan-Calvopina et al., 2016, Seisenberger et al., 2012). Those genes are therefore named "methylation-sensitive genes". Active demethylation at promoter regions may be required for expression of those germline-specific genes and may also affect other epigenetic modifications (Tang et al., 2015).

Other key features, at both transcriptional and epigenetic level, are reported to be different between mouse and hPGC development. In mPGCs, DNMT1, in the absence of Uhrf1, is suggested to maintain methylation on imprint control regions, X-linked genes and intracisternal A particles (IAPs) during the replication-dependent passive demethylation (i.e. Stage I demethylation), which might be similar to the function of DNMT1 and DNMT1o in DNA methylation maintenance in preimplantation embryo and oocyte. Those specific loci, except IAP, are demethylated via TET-active demethylation (i.e. Stage II demethylation) after reaching gonadal ridge. Conditional deletion of DNMT1 in PGC lead to abrogation of the two-stage DNA demethylation, precocious germline differentiation and fewer E13.5 PGCs (Hargan-Calvopina et al., 2016). This temporal and mechanistic separation of two stages of DNA demethylation in mPGC may not be conserved in hPGCs. The timing of imprint erasure are relatively earlier in hPGCs of which could start in migratory stage (before Wk4), compared to mPGCs in which the phenomena are observed in early gonadal PGCs (after E8.5) (von Meyenn and Reik, 2015, Guo et al., 2015, Tang et al., 2015, Chuva de Sousa Lopes et al., 2008). This

observation is further supported by the upregulation of TET1 and detection of 5hmC in Wk4 hPGCs (Guo et al., 2015, Tang et al., 2015). Partial erasure of 5mC coupled with increased level of 5hmC at H19 and GNAS ICRs are also reported in D4 hPGCLCs. Although another study reported that observant demethylation at H19 promoter region is only detected after Wk9 hPGCs, the bisulfite-PCR used in that study only detected methylation status within a 200bp region (Eguizabal et al., 2016). Still, inaccessibility of human embryos at earlier stages prohibit investigation of *in vivo* demethylation during hPGC specification and early migration, which emphasizes the necessity of a reliable model animal and robust *in vitro* induce system for studying 5mC and potentially 5hmC dynamics in early PGC development.

Epigenetic reprogramming in PGCs of other species has been recently reported. Global reduction of 5mC and H3K9me2, as well as repression of their epigenetic modifiers such as Uhrf1 and EHMT2/G9a, is evident in newly specified (E14) and early migratory (E16) pig PGCs compared to surrounding somatic cells, which indicates an earlier initiation of epigenetic reprogramming compared to in mouse but consistent with the timing in human. Interestingly, along with the reduction of 5mC, TET1 and 5hmC level are significantly enriched in pre- and early migratory pig PGCs, which is consistent with the observation in Wk4 hPGCs and further indicates temporal and mechanistic difference of epigenetic reprogramming between rodents and non-rodent mammalian PGCs (Kobayashi et al., 2017). Similarly, 5mC and H3K9me2 is significantly lower in cyPGCs from as early as around E17 and E20, respectively (Sasaki et al., 2016). Given the similarity observed between human and porcine embryo at morphological, transcriptional and epigenetic level as well as the high accessibility of porcine embryos, pig could be a surrogate model for investigating epigenetic reprogramming in human early PGCs, when combined with *in vitro* mimics for human embryonic development (Kobayashi et al., 2017).

Although no conclusive proof that DNA methylation can be inherited across generations, there is evidence suggesting that DNA methylation at some genetic regions can escape epigenetic reprogramming in both preimplantation embryos and PGCs (Tang et al., 2015, Guo et al., 2015, Hackett et al., 2013, Seisenberger et al., 2012). These regions are termed ‘escapees’, of which most

are repeat-rich TEs. However, a group of these escapees is repeat-poor, single copy genes which are associated with metabolic and neurological disorders. Interestingly, those repeat-poor 'escapees' are enriched in gene bodies. In contrast to the strong link between repressed gene expression and promoter methylation, methylation at gene bodies seems to be correlated with active gene expression (Chen et al., 2016b, Jones, 2012, Schroeder et al., 2015).

Overall, redundant mechanisms have evolved to act in parallel for the profound erasure of epigenetic marks in early embryos and PGCs, yet several mechanisms also exist to protect methylation at specific genomic regions. Those demethylation and methylation-protection mechanisms differ in details among mammalian species, suggesting a recent and presumably ongoing evolutionary selection. Future research looking into details of these mechanisms in different mammalian species would present a comprehensive network for regulation of epigenetic states in development and inheritance which could have promising biomedical applications and demonstrate how the cycle of life and developmental pattern might have evolved in mammals.

1.2.1.2 Functions and regulation of TET proteins and 5hmC

The non-reciprocal pattern between 5mC and 5hmC, and the dynamics observed during late oogenesis and early embryo development suggest additional roles of 5hmC other than just being a transient status during demethylation (Masala et al., 2017). 5hmC is a relatively stable and abundant DNA modification (Zhao et al., 2014), and have different binding properties compared to 5mC. Various 5hmC 'readers' have been identified with multiple high-throughput approaches, and many of them have been determined as transcriptional factors and DNA repair components (Mellen et al., 2012, Yildirim et al., 2011, Spruijt et al., 2013, Zhu et al., 2016, Mohni et al., 2019) (Figure 1-8).

Extensive research in mESCs and several other cell lines have reported potential activities and roles of TET proteins and 5hmC (Wu and Zhang, 2011, Xu et al., 2011, Cimmino and Aifantis, 2017). TET enzymes bind preferentially to promoters and gene bodies but also enhancers with affinity that positively correlates with CpG density (Williams et al., 2011, Stroud et al.,

2011). TET1, 2 and 3 share an evolutionally conserved Cysteine-rich and typical double-stranded beta helix fold of 2OG-Fe(II) oxygenases, but TET1 and 3 contain a DNA binding CxxC domain, which TET2 lacks. Evidences for the functional redundancy and specificity of TET enzymes are also controversial, suggesting TET activities are likely to be highly context-dependent (Huang et al., 2014, Hon et al., 2014). Interactions of TET protein with pluripotent genes, such as Nanog (Costa et al., 2013, Pantier et al., 2020), Oct4 (Piccolo et al., 2013), and Lin28A (Tan and Yeo, 2016, Zeng et al., 2016, Jiang and Baltimore, 2016) have also been reported in different systems. It is tempting to propose that those interactions also occur in early pig PGCs and play a role during pig PGC development.

Unlike the strong association of 5mC with gene silencing, the correlation between 5hmC and gene expression appears to be much more diverse. Also, genomic 5hmC pattern is shown to be highly dynamic, which is suggested to be related to tissue-specific biological processes (Hendrich et al., 2001, Ponnaluri et al., 2017, Lin et al., 2017, Nestor et al., 2012). Although 5hmC enrichment on gene bodies has been mostly suggested to link with active gene expression in various contexts (Guo et al., 2016, Lin et al., 2017, Wu and Zhang, 2011, Ponnaluri et al., 2017), 5hmC on promoters and enhancers are reported to be correlated to active (Deplus et al., 2013, Ficz et al., 2011, Stroud et al., 2011) or repressed genes (Choi et al., 2014, Williams et al., 2011, Wu and Zhang, 2011, Zhang et al., 2016e, Kim et al., 2014) by different studies.

TET activities and the genomic content of 5hmC could also contribute to the regulation of expression of specific genes during development and differentiation by modulating the open or closed state of the chromatin (Ruzov et al., 2011). In foetal germ cells, genes with intermediate 5hmC levels at promoter regions are more likely to have higher chromatin accessibility and also have higher expression levels (Guo et al., 2016). TET activity on insulators may also regulate gene expression by altering chromosomal architecture. For example, TET1/2 are suggested global regulators of CTCF-dependent splicing, as the oxidation of 5mC to 5hmC abolish the inhibitory role of 5mC on CTCF binding (Marina and Oberdoerffer, 2016).

5hmC and TET activities also correlate with other epigenetic modifiers and therefore could contribute to a tissue-specific epigenetic profile. For example,

the correlation between 5hmC and bivalency, a poised epigenetic state characteristic of pluripotent cells, has been reported by multiple studies (Pastor et al., 2011, Neri et al., 2013a, Gao et al., 2013, Kong et al., 2016, Verma et al., 2018). The role of TETs, especially around TSSs, can be independent of its catalytic function (i.e. the exact level of 5mC or 5hmC) (Williams et al., 2011), and more related to its binding at CpG-rich region and interaction/recruitment with other proteins such as OGT (Deplus et al., 2013, Vella et al., 2013, Chen et al., 2013) and PRC2 complex (Xu et al., 2011, Vella et al., 2013, Wu et al., 2011, Neri et al., 2013a). This may partially explain the finding of both active and repressive role of 5hmC. It was unclear until recent whether TETs-mediated active demethylation and 5hmC have site-specific roles on transcriptional regulation during PGC development. Recent studies in gonadal mPGCs and gonocytes suggest TET1 targets germline reprogramming-responsive (GRR) genes (Hill et al., 2018), imprinting control regions (ICRs) and meiotic genes (Hargan-Calvopina et al., 2016). TET1-mediated DNA demethylation at those sites provides protection from de novo DNA methylation and maintain the demethylated states, which facilitates timely expression of these germline and meiotic genes. However, 5hmC itself was suggested not essential for activating those genes (Hargan-Calvopina et al., 2016) and instead. Given the different tempo of DNA demethylation observed between mouse and non-rodent PGCs, it is possible TET and 5hmC have different functions in non-rodent PGCs at earlier stages.

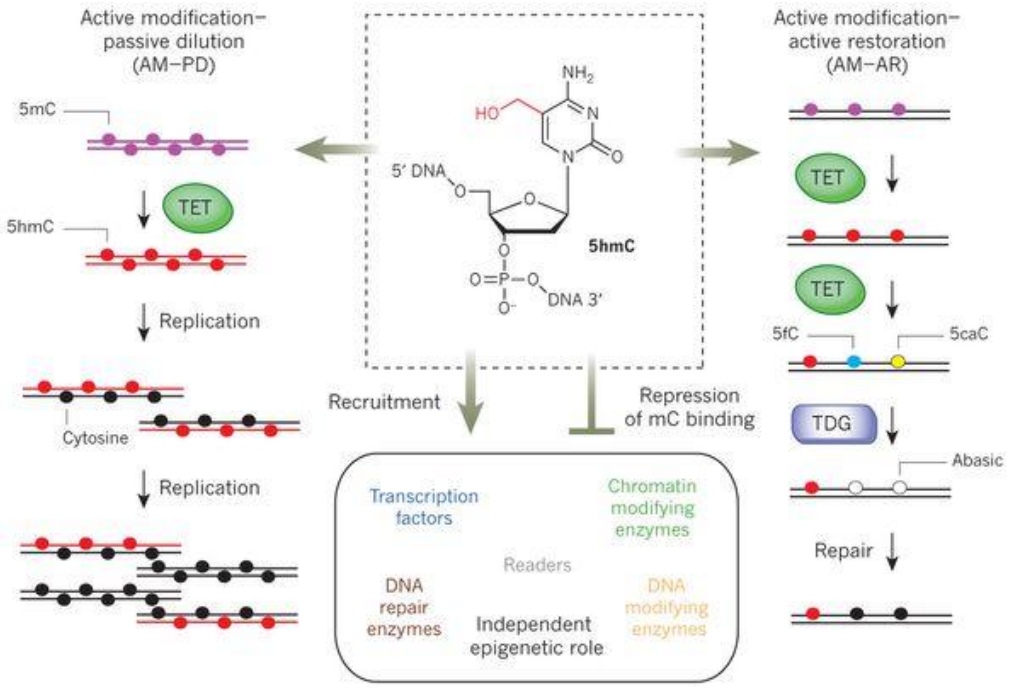


Figure 1-8 TET activity and removal of TET-oxidative products (Kohli and Zhang, 2013). TET proteins oxidize 5mC into 5hmC (i.e. Active modification), which can be removed during DNA replication (i.e. passive dilution), or further oxidized by TET into 5fC and 5caC, which is removed by DNA repair mechanisms (i.e. Active restoration). 5hmC may be a functional DNA modification instead of a by-product during DNA demethylation.

1.2.2 **Histone variants and modification during mammalian development**

1.2.2.1 H3K9me2/3

H3K9me2/3, like other epigenetic markers, form in a cell type-specific manner, and are required for securing cell fate and maintaining cell identity (Becker et al., 2016). In general, methylation of H3K9 is considered a repressive mark. H3K9me2/3 is associated with heterochromatin, as H3K9me2/3 provides a docking site for heterochromatin protein 1 (HP1) which are required for formation of heterochromatin (Canzio et al., 2011). SETDB1 (Liu et al., 2014), SUV39H1 and SUV39H2 contribute to both H3K9me2 and H3K9me3, although the binding of SUV391 to H3K9me2 is weaker (Wang et al., 2012).

H3K9me2 can also be catalysed by GLP (EHMT1) and G9a (EHMT2), which also contribute to H3K9me1 (Tachibana et al., 2002, Tachibana et al., 2005). The functional difference among H3K9me1, H3K9me2, and H3K9me3 remain largely unclear, although the dynamics and distribution of H3K9me2 and H3K9me3 are shown different in *C.elegans* (Bessler et al., 2010, Zeller et al., 2016, McMurchy et al., 2017) and in mESCs (Walter et al., 2016). A recent study in fission yeast has reported major functional differences between H3K9me2 and H3K9me3. Although both histone marks are associated with formation of heterochromatin, H3K9me2 is found preceding the occurrence of H3K9me3, co-existing with transcriptional active marks and related with expression of siRNA and co-transcriptional gene silencing (CTGS), whereas H3K9me3 is essential for recruitment of downstream factors and complexes that are required for transcriptional gene silencing (TGS). Also, H3K9me3 is required for inheritance of RNAi-independent H3K9 methylation and gene silencing at pericentromeric repeats, which could due to a higher affinity of H3K9 methylation 'writer' for H3K9me3 compared to H3K9me2 that allows stable binding to parentally inherited histone (Jih et al., 2017).

H3K9me3 at heterochromatin regions, retrotransposons and 'escapees' is suggested to be essential for genome stability and resistance of DNA demethylation, respectively, during global erasure of DNA methylation. In early mammalian embryos, H3K9me2/3 protects DNA from TET-mediate active demethylation via recruitment of STELLA/PGC7 (Wossidlo et al., 2011, Nakamura et al., 2012, Nakamura et al., 2007). Methylation on H3K9 may also be critical for the profound de novo DNA methylation after epigenetic reprogramming, as interactions of H3K9-methylated heterochromatin and contributors of H3K9-methylation with DNMT3a/b and Uhrf1 have been reported on several occasions (Liu et al., 2013, Rose and Klose, 2014, Rothbart et al., 2013, Zhao et al., 2016). Remodelling of H3K9 methylation are also essential for cell fate transition in early mouse embryo (Dodge et al., 2004, Tachibana et al., 2005, Tachibana et al., 2002, Wang et al., 2010, Zylicz et al., 2015) and in pig (Park et al., 2011). Most interestingly, in contrast to the re-establishment of constitutive heterochromatin by PRC1/2 in mouse preimplantation embryo, it is indicated that H3K9 methylation in constitutive heterochromatin is retained in human sperm, which is recognized by maternal

chromatin modifiers in H3K9/HP1 pathway and propagated over the embryonic cleavage divisions. This observation indicates H3K9 methylation in sperm contribute to the formation of constitutive heterochromatin early human embryo, and therefore is potential 'carriers' for paternal epigenetic information (van de Werken et al., 2014).

The distribution and remodelling of H3K9me2 and H3K9me3 are also shown to be slightly different in mammalian PGC development (Tang et al., 2015, Seki et al., 2007, Hajkova et al., 2008). Both H3K9me2/3 undergo global reduction in mouse and hPGCs. However, H3K9me2 is still detectable in the hPGCs whereas it is extensively depleted in mPGCs (Tang et al., 2015). H3K9me3 is maintained in DAPI-dense peri-centric heterochromatin foci and 'escapee' regions, which is observed as a 'punctate pattern' by immunofluorescence. The exact function of H3K9me2/3 in PGCs remains to be explored.

1.2.2.2 H3K27me3

H3K27me3 is a repressive mark for gene expression and a hallmark of facultative heterochromatin in various tissues and cell lines. It is tightly linked to inactive gene promoters and suggested to maintain transcriptional repression. H3K27me3 is catalysed by Polycomb Repressive Complex 2 (PRC2), of which the main components are Ezh1/2, Eed, and Suz12. Those PRC2 components are ubiquitously expressed. As the core components of PRC2 complex does not contain any DNA-binding domains itself, the mechanisms of its recruitment to genomic loci is poorly understood. Still, several recruitment factors, such as Jarid2 and PCL, have been suggested to involve in the targeting of PRC2 (Holoch and Margueron, 2017). The recruitment of PRC2 to genomic loci is likely to be affected by local chromatin environment, which includes GC content (Mendenhall et al., 2010), DNA methylation and protein binding profiles. For example, disruption of normal DNA methylation leads to abnormal PRC2 binding and H3K27me3 profiles in mammalian adult cells (Lynch et al., 2012, Reddington et al., 2013). However, evidence about the relationship between 5mC and H3K27me3 seem to be controversial. H3K27me3 and DNA methylation co-exist at CpG-poor promoter regions

(Statham et al., 2012, Brinkman et al., 2012). A direct interaction between EZH2, the catalytic component of PRC2, and DNMT3A/B has also been demonstrated in mammalian cell lines (Vire et al., 2006, Rush et al., 2009). However, DNA methylated chromatin inhibit binding of PRC2 *in vitro* (Bartke et al., 2010, Wu et al., 2010). Negative correlation of H3K27me3 with DNA methylation has also been widely observed CpG-rich regions (Lister et al., 2009, Meissner et al., 2008, Hawkins et al., 2010, Lindroth et al., 2008). TET-mediated demethylation may also be correlated with PRC2 activity, as the interaction of TET1 and PRC2 has been reported in mESCs (Neri et al., 2013a). The reprogramming of H3K27me3 in mPGC occurs from E8.25 to E13.5 (Kagiwada et al., 2013, Hajkova et al., 2008, Seki et al., 2005, Seki et al., 2007), when mPGCs start to show enrichment of H3K27me3 compared to surrounding somatic cells. It was shown that the core component of PRC2, Eed, is only transiently (E11.5-12.5) enriched in mPGCs, while the other two components Ezh2 and Suz12 are present between E11.5-15.5. E11.5-13.5 mPGCs also show peripheral nuclear enrichment of H3K27me3, which is dependent on EZH1/2 and coincides with the final phase of DNA demethylation and the lowest levels of DNA methylation in germ cells. By E15.5, H3K27me3 was again evenly distributed in mPGCs. However, enrichment of H3K27me3 is not necessary for mPGC development (Prokopuk et al., 2017). Instead, ChIP-seq in E13.5 mPGC suggest that H3K27me3 may be important to keep retrotransposons repressed and to protect genomic integrity, as H3K27me3-deposit regions in mPGCs are enriched for retrotransposon repeats (Ng et al., 2013). There are also contradictory evidences suggesting that H3k27me3 is lost in PGCs at E11.5 and re-gained between E12.5 and E13.5 (Hajkova et al., 2008), and the demethylase KDM6A is reported to be responsible for the transient loss and is important for development of mPGCs (Mansour et al., 2012). Dynamics of H3K27me3 in hPGC is slightly different compared to in mPGCs (von Meyenn and Reik, 2015). Unlike the persistent enrichment of H3K27me3 in mPGCs, H3K27me3 is enriched in migratory hPGCs but depleted from Wk7 onward (Tang et al., 2015). It is worth noting that the reprogramming of H3K27me3 might also be important for germline competency, as knockout of PRC2 components leads to de-repression of BMP signalling and spontaneous differentiation toward a

meso-endoderm fate (Shan et al., 2017). H3K27me3 is also viewed as a marker for X-chromosome inactivation (Plath et al., 2003, Mak et al., 2002). Difference of H3K27me3 remodelling in human and mouse PGCs may also affect other chromatin remodelling events, such as X chromosome reactivation (XCR, see section 1.2.3) is regulated in those two species (von Meyenn and Reik, 2015). However, details of the difference are unknown due to lack of locus-specific data in human or pre-gonadal mPGCs.

In preimplantation mouse embryo, a massive loss of H3K27me3 is observed in zygote and the two-cell-stage, which is coincident with the zygotic genome activation (Lee et al., 2014b, Zheng et al., 2016). A highly dynamic remodelling of H3K27me3 also occurs during the transition from the morula to the ICM/TE. Function annotation of the H3K27me3-marked promoters in preimplantation embryos reveals enrichment for further development, such as "neuron differentiation" (Liu et al., 2016). Interestingly, a recent study indicated H3K27me3, but not H3K9me3, function as an imprinting mechanism independent of DNA methylation, which is observed on maternal allele and maintain paternal allele-specific chromatin accessibility and gene expression in early preimplantation embryo. However, this imprinting mechanism start to diminish in ICM and completely lost in epiblast. Only a small portion of this imprinting is maintained in extra-embryonic tissue (Inoue et al., 2017). It is possible that H3K27me3-dependent imprinting function as alternative to DNA methylation-dependent imprinting, especially during global erasure of DNA methylation. Therefore, it would be interesting to see if this imprinting mechanism also occur during DNA demethylation at imprinted regions in PGCs.

1.2.2.3 H3K4me2/3

Unlike the repressive H3K9me2/3 and H3K27me3, H3K4me2/3 is considered an active mark for gene expression (Schuettengruber et al., 2007). H3K4me2/3 is established by SET1A/B complex and Mll complex. ChIP-seq often reveals H3K4me2/3 as narrow peaks (~100bp) around transcriptional start site (TSS), which are mutually exclusive with DNA methylation (Weber et al., 2007). The

observation that Dnmt3L is able to bind to non-methylated H3K4 but cannot bind to H3K4me3 suggests that H3K4 methylation may also play a role in blocking de novo DNA methylation at some genomic loci. Consistent with the requirement of Dnmt3L in de novo DNA methylation and establishing genomic imprinting in oocyte (Henckel et al., 2009), an H3K4me2 demethylase, KDM1B, is also shown to be required to establish maternal genomic imprinting (Ciccone et al., 2009). However, recent studies report two types of broad H3K4me3 in oocytes and early mouse embryo (Zhang et al., 2016a, Dahl et al., 2016). One type is found around TSSs and positively correlated with transcriptional level, whereas the other type is found at regions distal to TSSs. Those broad H3K4me3 regions overlap almost exclusively with partially methylated DNA domains in oocyte and are inherited in preimplantation embryo, but soon erased in the late two-cell stage and replaced by canonical narrow H3K4me3 (Zhang et al., 2016a). Interestingly, defects in genome silencing are observed in full-grown oocytes with overexpressed KDM5B, an H3K4me3 demethylase, indicating a potential role of broad H3K4me3 regions in gene silencing (Zhang et al., 2016a). KDM5A and KDM5B is also essential for mammalian preimplantation embryo (Dahl et al., 2016). Knockdown of KDM5B is embryonic-lethal in pig, which also lead to upregulation of other epigenetic modifiers such as KDM6A and TET proteins (Huang et al., 2015). Another demethylase for H3K4me3, KDM2B, has been reported to spermatogenesis in mice (Ozawa et al., 2016).

H3K4me3 may also be related to 5hmC and TET-mediated demethylation, as it is shown that Tet1 is positively correlated with H3K4me3 at CpG-rich promoters in mESCs (Xu et al., 2011). The interaction between TET proteins and OGT (O-GlcNac transferase) which glycosylates HCF1, a component of the Set1a/b and MLL1/2 complexes, suggesting that this may function as another means to recruit H3K4 methylation to CpG-rich regions(Vella et al., 2013, Deplus et al., 2013, Clouaire et al., 2012).

Despite of the profiles and potential functions of H3K4me3 reported in mammalian embryo and mature germline, little is known about H3K4me3 in PGCs IF shows that H3K4 methylation enriches in mPGCs from E8.5 (Hajkova et al., 2008), whereas the level in hPGCs are largely similar to surrounding somatic cells (Eguizabal et al., 2016). . Specific H3K4me3 peaks

are found around Dnmt3b promoter in mouse EpiLCs, promoter of T in d2 mPGCLCs and Blimp1 promoter in d2 and d6 mPGCLCs, which are consistent with their activation.

1.2.2.4 Bivalency at promoters

Bivalency, or bivalent chromatin domains, has been defined as individual nucleosome physically co-occupied by both activate histone modification H3K4me2/3 and repressive histone modification H3K27me3 (Bernstein et al., 2006). The concept of bivalency is biochemically unusual as the mechanisms for establishing H3K4me3 and H3K27me3 were considered mutually exclusive (Young et al., 2009, Schmitges et al., 2011) (Kim et al., 2013a, Kinkley et al., 2016). However, unique arrangements at a small set of genomic regions cannot be ruled out (Voigt et al., 2012, Schmitges et al., 2011). Also, the co-existence of H3K4me3 and H3K27me3 at same genomic regions has been proved with reChIP-seq (aka Sequential ChIP-seq) technique (Bramswig et al., 2013)

Bivalent regions were first identified in mouse ESCs (Bernstein et al., 2006), and are often revealed by ChIP-Seq as narrow H3K4me3 peaks with broad H3K27me3 signal at promoter regions. Notice that bivalency also occurs on enhancer regions but is defined as co-existence of H3K27ac and extensive DNA methylation (Charlet et al., 2016), however, enhancer regions are in general not annotated in pig genome, and therefore will not be discussed here. Bivalency at promoter regions is suggested to function as a specific mark which keeps genes poised in an “activable” state, allowing silenced genes to respond rapidly to environmental signals, such as developmental cues. The early analyses between ES cells and differentiated cell types indicate that bivalent domains are characteristic of pluripotent cells and they mark the poised developmental genes which are silenced but could be rapidly activated (Mikkelsen et al., 2007). Although later studies reveal the existence of bivalent promoters in a wide range of cells (Mohn et al., 2008, Bhandare et al., 2010, Lee et al., 2014c, Kinkley et al., 2016, Triantaphyllopoulos et al., 2016), bivalency in majority of studies, regardless of the cell or tissue examined, shows enrichment in core development-related genes.

Bivalent domains have been identified in mouse gonadal PGC (Lesch et al., 2013, Sachs et al., 2013), PGCLCs (Kurimoto et al., 2015), adult testes of a few mammalian species (Hammoud et al., 2014) and human spermatozoa (Brykczynska et al., 2010, Lesch et al., 2016). Sachs et al. showed that bivalent domains largely occur at transcriptional start sites (TSSs) in both mESCs and E11.5 mPGCs but not in the surrounding somatic cells. Bivalent genes in mESCs and E11.5 mPGC show a significant overlap and are largely enriched in developmental regulators. Lesch et al. indicate a cluster of 147 germline-poised genes, which are bivalent genes highly conserved in E12.5 and E14.5 mPGCs and pachytene spermatocytes and post-meiotic round spermatids. Those 147 genes are conserved between sex, and function as developmental regulator in early embryos. Bivalency identification in male germ cells across five mammalian species, reported by Lesch's group, further indicates a deep conservation of poised gene in mammalian male germ cells, which implies a functional role for the poised state in germline, presumably related to somatic development after fertilization. Those core poised genes are suggested to be 'epigenetically privileged'(Lesch et al., 2016), and are interesting targets for epigenetic inheritance study.

Due to limited number of *in vivo* PGCs at early stages, locus-specific information of histone modification is currently only available for mouse gonadal PGCs and mouse PGCLCs. During the induction of mPGCLCs *in vitro*, the profile of bivalency is highly dynamic and correspond to the identity of the cells. EpiLCs contain the largest number of bivalent genes and are characterized by low H3K27me3, indicating a state primed for differentiation. In d2 PGCLCs, a considerable proportion of bivalent genes in EpiLCs become H3K27me3 only and presumably repressed. Those newly repressed genes are highly enriched in 'neuron differentiation', indicating loss of potential in somatic programming such as 'neuron differentiation' and restriction to germline lineage. Interestingly, a re-acquisition of bivalency due to regain of H3K4me3 is observed in d6 PGCLCs, which shows 58% of overlap with E11.5 mPGCs (Kurimoto et al., 2015). Given the large overlap of bivalent domain in mESCs and E11.5 mPGCs, it is tempting to speculate that the re-acquisition of bivalency in early germline is linked to its ability to reacquire totipotency after fertilization. Therefore, dynamics of bivalent promoters and

underlying mechanisms in PGCs and preimplantation embryo may help us understand the continuation of mammalian life cycle

1.2.2.5 macroH2A

Similar to histone post-translational modification (PTM), histone variants, which are non-allelic counterparts of canonical histones, also regulate chromatin functions and participate in various biological processes, such as transcriptional regulation and DNA damage repair (Santoro and Dulac, 2015, Buschbeck and Hake, 2017, Banaszynski et al., 2010). Replacement of canonical histones with histone variants endows chromatin with unique biochemical and structural features, and therefore leads to unique chromatin landscapes and transcriptional program. Unlike canonical histones which are generally produced and deposited uniformly on DNA during replication (i.e. replication-coupled), histone variants can be incorporated in a locus-specific manner throughout cell cycle. This replication-independent deposition relies on specific machineries and chaperones (Burgess and Zhang, 2013), which enables histone variants to have spatiotemporal-specific distribution and cellular functions.

Numerous histone variants have been discovered in eukaryotes. Here I focus on macroH2A, a highly conserved histone variant in vertebrates (Rivera-Casas et al., 2016). MacroH2A contains an H2A-resembling domain, followed by a large non-histone macro domain on the C-terminus. Three isoforms, macroH2A1.1, macroH2A1.2 (Fried, 1992, Pehrson et al., 1997) and macroH2A2 (Chadwick and Willard, 2001, Costanzi and Pehrson, 2001) are identified in mammals. MacroH2A1.1 and macroH2A1.2 are splice variants encoded by H2AFY, and macroH2A2 is encoded by a different gene H2AFY2. The three isoforms show redundant yet also specific functions in different contexts (Gonzalez-Munoz et al., 2019, Creppe et al., 2012, Buschbeck et al., 2009). The regulation of macroH2A expression and deposition is still largely unknown. The chromatin remodeller ATRX is suggested as a negative regulator of macroH2A on telomeres, and a few chromatin chaperons possibly promote macroH2A deposition during DNA damage repair (Kim et al., 2018b, Mehrotra et al., 2011).

The large macro domain in MacroH2A is about twice the size of canonical H2A, which can largely change nucleosome properties and affect local transcription when MacroH2A is incorporated. Conventionally, macroH2A has been suggested as a repressive mark, due to its enrichment on the inactive X chromosome (Costanzi and Pehrson, 1998) and constitutive heterochromatin (Douet et al., 2017), as well as co-occupation with other repressive marks, especially H3K27me3 (Gaspar-Maia et al., 2013, Buschbeck et al., 2009). MacroH2A has also shown interaction with histone deacetylase and PRC2 complex, which contribute to the inactive state of specific genes (Chakravarthy et al., 2005, Buschbeck et al., 2009, Kim et al., 2013b, Kim et al., 2018a, Wan et al., 2017a). Deposition of macroH2A also associate with DNA methylation, as macroH2A is detected at the methylated CGI of several imprinted genes, including the imprinting control regions (ICRs) of Xist, Peg3, H19/Igf2, Gtl2/Dlk1 and Gnas (Choo et al., 2006, Changolkar et al., 2007). The ability of macroH2A to promote chromatin condensation also enables its role in DNA repair (Kozlowski et al., 2018, Kim et al., 2018b, Xu et al., 2012). However, a few studies have demonstrated macroH2A can play a positive role in a context- and isoform-dependent way (Gamble et al., 2010, Pehrson et al., 2014, Dell'Orso et al., 2016, Ruiz and Gamble, 2018). Different isoforms of macroH2A, namely macroH2A1.1 and macroH2A1.2, are found preferentially associated with repressive or active marks, respectively (Podrini et al., 2015, Wan et al., 2017a). Overexpression of either isoform leads to change in bivalent promoters, suggesting that different isoforms of macroH2A may regulate gene expression through directing the bivalent promoters to either open or close status (Podrini et al., 2015, Barrero et al., 2013b).

Not only affecting gene expression in a locus-specific way, one specific macroH2A isoform, macroH2A1.1, also has a more global role through interaction with PARP1 (Hurtado-Bagès et al., 2018, Posavec Marjanović et al., 2017, Hodge et al., 2018). The binding activity of MacroH2A1.1 towards ADP-ribose impedes basal PARP1 activity (Ladurner, 2003, Kustatscher et al., 2005), which is major consumer of NAD+. Thus, gain of macroH2A1.1 can limit nuclear NAD+ consumption and regulates mitochondrial respiration. It is also shown that macroH2A1.1 is rapidly recruited to PARP1 activation sites during DNA damage response, leading to a transient compaction of chromatin

and mediating recruitment of other repair components for non-homologous end joining (NHEJ) (Timinszky et al., 2009, Xu et al., 2012).

Expression and deposition of macroH2A is dynamically regulated during embryonic development and cell reprogramming (Figure 1-9). During mouse embryonic development, macroH2A is detected in mature oocytes, but undetectable from pronuclei to eight-cell stage. From the morula stage, macroH2A reappears and persists to the blastocyst stage (Chang et al., 2005). Most interestingly, macroH2A has been suggested as the barrier for reprogramming to or exit from pluripotency. Differentiated cells show higher level of macroH2A compared to pluripotency cells, and knockdown of macroH2A in mESC has shown limited capacity of differentiation (Creppe et al., 2012). MacroH2A may also obstruct reactivation of pluripotency genes, which reduces the efficiency of induced pluripotent stem cells (iPSCs) (Pasque et al., 2012, Gaspar-Maia et al., 2013, Pasque et al., 2011, Barrero et al., 2013a). Nevertheless, mice lacking all three isoforms of macroH2A are viable and fertile despite showing metabolic defects and reduced reproductive efficiency (Gaspar-Maia et al., 2013, Pehrson et al., 2014, Sheedfar et al., 2015, Boulard et al., 2010), suggesting that macroH2A may function as an additional layer to secure cell identity but is not required for the initiation of cell fate transition during embryonic development. Apart from in mature oocytes, macroH2A is also detected in gonadal mPGC (Nesterova et al., 2002, Hajkova et al., 2008), as well as on sex chromosome during male and female meiosis (Turner et al., 2001, Richler et al., 2000, Hoyer-Fender et al., 2000). However, macroH2A2 is largely reduced in Wk4-9 hPGC compared to surrounding somatic cells (Tang et al., 2015). It remains largely unknown the regulation and function of macroH2A during mammalian germ cell development.

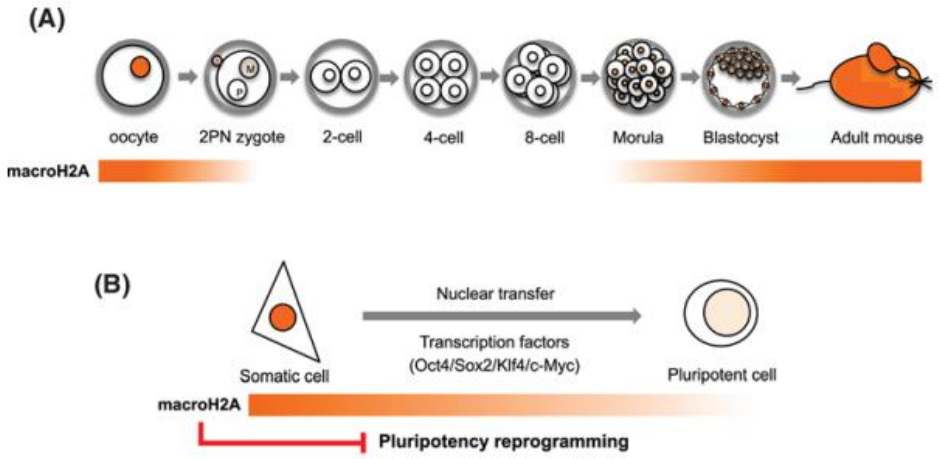


Figure 1-9 Dynamic changes of MacroH2A during development and pluripotency reprogramming (Sun and Bernstein, 2019). MacroH2A is enriched in determined cell lineages, including mature oocyte, but becomes depleted in pluripotent cells in vivo and during reprogramming to pluripotency.

1.2.3 X chromosome status during mammalian development

Mammalian sex chromosomes (X and Y) contain highly divergent gene content (Soh et al., 2014). As X chromosome dosage, especially of non sex-specific genes, is suggested to be crucial for mammalian development and fitness (Schulz and Heard, 2013, Deng et al., 2014), X chromosome inactivation (XCI) has evolved to transcriptionally silence one out of two X chromosome in females and to achieve dosage compensation between female (XX) and male (XY) individuals (Lyon, 1961, Ohno, 2013). The reversal process of XCI, X chromosome reactivation (XCR), occurs in female preimplantation embryo and primordial germ cells (PGCs), is also critical for mammalian development (Pasque and Plath, 2015, Cantone and Fisher, 2017). The transition of X chromosome status involves complex regulation at multiple epigenetic levels, which eventually lead to the structural and functional alteration of one whole chromosome. This makes XCI and XCR paradigms for reprogramming progression and good tools for studying mechanisms of epigenetic reprogramming.

1.2.3.1 X chromosome status during the mouse life cycle

Mammalian XCI is most extensively investigated in mice, and two waves of XCI initiate during early development (Figure 1-10). An imprinted XCI of the paternal X chromosome (Xp) initiates between 2-4 cell stages (Okamoto et al., 2005, Okamoto et al., 2004, Borensztein et al., 2017b). Reactivation of Xp is detected in the inner cell mass (ICM), followed by a random (that is, without parental-of-origin bias) and heritable XCI in epiblast (Monk and Harper, 1979), while the Xp is maintained inactive in extra-embryonic tissue (Mak et al., 2004). Both imprinted and random XCI begin with the expression and accumulation of the long non-coding RNA (lncRNA) Xist (X-inactive specific transcript) on the chromosome that will be silenced (Brown et al., 1991). The inactive state after random XCI is further reinforced by removal of RNA polymerase II (RNA Pol II) and active histone marks, deposition of repressive histone marks (H3K27me3, H3K9me3(Okamoto et al., 2004, Heard et al., 2001), H2AK119ub (Escamilla-Del-Arenal et al., 2011), macroH2A (Costanzi et al., 2000) and DNA methylation, as well as by changes in chromosomal conformation (Deng et al., 2015, Chaumeil et al., 2006).

XCR is considered the reversal of XCI. However, the underlying mechanism of XCR is not necessarily a direct reversal of XCI mechanisms, and it is much less known compared to XCI mechanisms. In mice, XCR occurs in female preimplantation embryo and primordial germ cells (PGCs). The reactivation of Xp in female mouse ICM leads to a short presence of two active X chromosomes, which is considered as a hallmark of naïve pluripotency and a striking example of chromatin remodelling during cell fate reprogramming. In vitro reprogramming of somatic cells to pluripotent stem cells is also coupled with XCR (Tada et al., 2001). In contrast to the rapid XCR of imprinted XCI in blastocyst, XCR in PGCs proceeds in an asynchronous way that spans days and multiple cell divisions. Initiation of XCR, shown by the absence of Xist cloud, is observed in very few mPGCs at E7.0 (pre-migratory), and becomes more evident (16.5% of mPGCs) at migratory stage (E8.5). Majority of mPGCs are Xist-negative and H3k27me3-negative after colonizing the genital ridge. However, E14.5 mPGCs still contain mono-allelically expressed X-

linked genes, suggesting XCR is not entirely complete even by this stage (Sugimoto and Abe, 2007, Chuva de Sousa Lopes et al., 2008, de Napoles et al., 2007). The active state of both X chromosome persists in matured oocytes (Epstein, 1969, Fukuda et al., 2015).

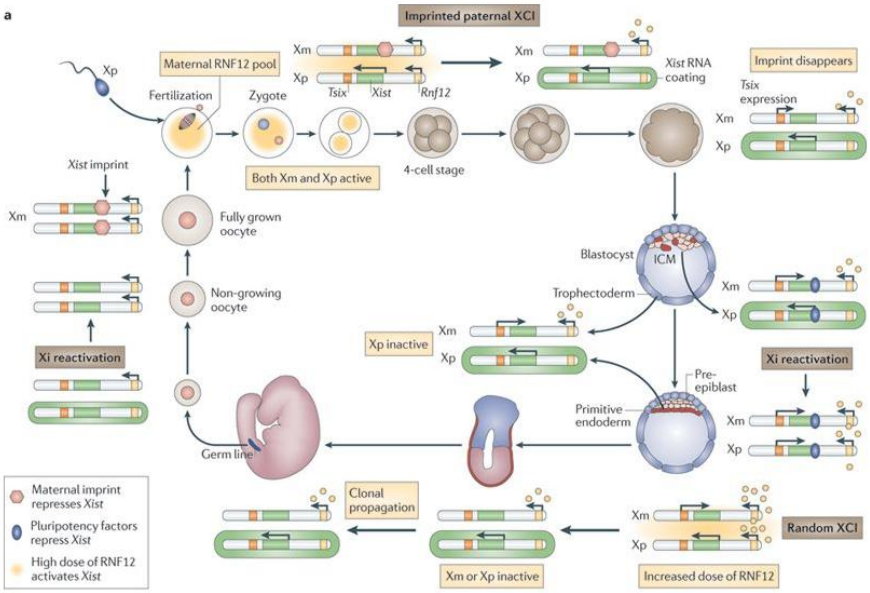


Figure 1-10 X chromosome status in mouse life cycle (Augui et al., 2011). In mice, preferential inactivation of the paternal X chromosome (Xp), i.e. imprinted paternal XCI, occurs at 4-cell stage. However, note that the imprint is actually on maternal X chromosomes (Xm), suggested by experiments with parthenogenetic and gynogenetic embryos. This imprinted paternal XCI is not observed in other non-rodent species such as human and rabbit (Okamoto et al., 2011). Established imprinted XCI is maintained in extra-embryonic lineages but reversed in the inner cell mass (ICM) of the blastocyst, leading to a transient period of two active X chromosome (i.e. Xi reactivation). In ICM, random inactivation occurs and the inactive X chromosome (Xi) is maintained during cell division in somatic lineages. However, this Xi is reactivated in female mPGCs. Xist is essential for XCI in mice, and several Xist-interacting factors, such as RNF12, have been identified.

1.2.3.2 Key difference of XCI/XCR between rodent and other mammalian species

Studies in other mammals, including human, rabbits, and marsupials, reveal diverse strategies for the initiation and regulation of XCI, compared to rodents (Sado and Sakaguchi, 2013, Escamilla-Del-Arenal et al., 2011, Graves, 2015). For example, in human and preimplantation embryo, upregulation of XIST is detected on both X chromosomes in females (i.e. bi-allelic XIST expression) and the X chromosome in males, which persists for several days (Tachibana et al., 2012). A transient bi-allelic expression of Xist is also observed in a significant portion of cells from morula stage in rabbit and at a much less frequency in mouse (Patel et al., 2017, Okamoto et al., 2011, Mutzel et al., 2019). Despite the expression of *XIST,* chromosome-wide XCI in human is suggested not to be initiated by the blastocyst stage, and monoallelic expression of X-linked genes is detected by the late blastocyst stage of female embryo (Moreira de Mello et al., 2017, Patel et al., 2017). As species-specific differences have been reported for XCI, it is likely that the kinetic and regulation of XCR may also be diverse between species (Cantone, 2017, Tang et al., 2016, Pasque and Plath, 2015). For example, *XACT*, a weakly conserved, X-linked lncRNA, is suggested to antagonize *XIST* and prevent XCI on both active X chromosomes in pre-implantation human embryo (Vallot et al., 2017). Still, compared to XCI, mechanisms of XCR are much less well understood and largely restricted to mice. Therefore, non-rodent mammalian models are desirable for further understanding species-specific X chromosome dynamics. A human-resembling model is specifically important for studying XCR in hPGCs. Recent studies present controversial evidences about the kinetic of XCR in hPGCs (Tang et al., 2015, Guo et al., 2015, Vértesy et al., 2018). Due to inaccessibility of early hPGCs (pre-/early migratory) in vivo, conclusive evidences of XCR kinetic and mechanism in hPGCs are not available. Given that key differences between human and mPGC have been observed at epigenetic levels (von Meyenn et al., 2016, Tang et al., 2015, Irie et al., 2015), XCR in hPGCs may also be regulated differently compared to in mPGCs. Therefore, early pPGC provides a desirable system to investigate XCR in non-

rodent mammals, which may also shed light on the regulation of XCR in human and other non-rodent species.

1.2.4 ncRNA during mammalian development

Given the large number of variable small non-coding RNA (sncRNAs), it is not surprising that profiles and potential roles of most sncRNA in mammalian PGCs and preimplantation embryo are largely unknown. Still, several studies indicate extensive interplay between sncRNAs and other epigenetic markers, which are crucial for embryonic and germline development. For example, deletion of miRNA biogenesis enzyme DICER in mESCs leads to large loss of PRC2 core components EZH2 and SUZ12 at the promoter regions of bivalent genes, while total and chromatin associated EZH2 protein remained unchanged. 3′ UTRs of transcripts of the bivalent genes that lose EZH2 binding are highly enriched in sequence motifs for binding sites of several ESC miRNAs (Graham et al., 2016).

miRNA are crucial for PGC development. Poor PGC proliferation is observed in mice with PGC-specific deficiency of Dicer (Hayashi et al., 2008). Deficiency of mammal-specific miRNA cluster *mir-290–295* lead to defects in migrating primordial germ cells in both male and female mutant mice (Medeiros et al., 2011). Expression of miRNAs may also regulate epigenetic modifiers such as DNMT3a/b and contribute the epigenetic reprogramming in PGCs (Takada et al., 2009). It has also been suggested that a miR-372/let-7 axis regulates hPGCLC induction, and therefore might play a role in human *in vivo* PGC specification (Tran et al., 2016). More potential roles of miRNA in PGC development are yet to be explored (Bhin et al., 2015). PIWI-interacting RNAs (piRNAs) are germline-specific small non-coding RNAs and are reported to involve repression of TEs under the global erasure of DNA methylation in mPGCs via establishment of H3K9me3 (Pezic et al., 2014). A recent study in *C.elegans* further indicates the collaboration between piRNA and heterochromatin factors is crucial for silencing repetitive elements and preventing genotoxic stress in the germ line. Mutation of those heterochromatin factors causes sterility and show similar phenotype with disruption in relevant piRNA pathways (McMurchy et al., 2017).

sncRNAs and RNA modifications in mammals, but not DNA methylation or histone modification, have shown convincingly the participation in epigenetic inheritance (Rassoulzadegan et al., 2006, Chandler, 2007, Yuan et al., 2015, Chen et al., 2016d, Larriba and del Mazo, 2016). Notice that recent study in *C.elegans* has suggest a new mechanism of transgenerational epigenetic inheritance (TEI) of RNAi. Lipid droplets organelles, i.e germ granules, which contains small RNA and TEI-related factors, are found spatially and temporally ordered into lipid droplets assemblages (termed PZM). This spatially and temporally ordering of liquid droplet organelles is speculated to help organize and coordinate the complex RNA processing steps that underlie RNA-directed TEI (Wan et al., 2017b).

1.2.5 **Epigenetic inheritance**

1.2.5.1 Epigenetic response to environment in mammal

The concerning, rapid global rise in metabolic disorder is unlikely to be explained solely by genetic drift (Fuchsberger et al., 2016). Therefore, non-genetic environmental factors have been suggested to contribute to inheritance of disease risk. One of the most documented case of nutrient-related transgenerational effect is the Dutch Hunger Winter. Individuals exposed to famine in utero during the last two trimesters of gestation and their offspring show increased risk for metabolic disorders (Jimenez-Chillaron et al., 2015). Several similar cases have been documented in different time and regions of the world (Bock et al., 2010). A recent study further reports differential methylated regions related to a few metabolism/mental pathways in 8 grandchildren whose grandfathers/grandmothers have been exposed to famine/feast (Bygren et al., 2017). The epigenome of individuals is relatively dynamic and susceptible to environmental influences throughout early life and adulthood (Donkin et al., 2016, Hao et al., 2016, Jang and Serra, 2014, Huypens et al., 2016), allowing for rapid adaption to fluctuating environments and nutrition. Although most of those metastable epigenetic variations (i.e. epialleles) would be either neutral or deleterious, various studies in plants and invertebrates indicate transgenerational epigenetic inheritance has the potential

as a response to environmental challenges (Heard and Martienssen, 2014, Quarta et al., 2016).

However, evidence indicating that inheritable phenotypes, especially diet-induced transgenerational inheritance, have an epigenetic basis is still inadequate in mammals, and the mechanisms responsible remain largely unclear (Daxinger and Whitelaw, 2012, Lim and Brunet, 2013, Heard and Martienssen, 2014). In mammals, two efficient waves of global epigenetic erasure in early embryo and primordial germ cells, respectively, ensure removal of any potentially deleterious epigenetic marks. For example, deleterious effects of endocrine disruptors have been shown to be corrected through epigenetic reprogramming (Iqbal et al., 2015, Whitelaw, 2015). Also, the existence of physical barriers between somatic cells and germline cells (such as the blood-testis barrier) makes it difficult to get an insight into epigenetic inheritance. Therefore, a better understanding of how epigenetic marks may successfully evade epigenetic reprogramming and be transmitted across generations would be essential for understanding whether there is an 'epigenetic predisposition' to diseases (Vickers, 2014).

DNA methylation at the 'escapee' regions discovered in hPGCs represent exciting targets for the study of epigenetic inheritance. A few studies have reported alteration in DNA methylation in mammalian mature germ cells related to heritable phenotypes, which are induced by environmental factors including endocrine disruptors (Skinner, 2007, Skinner et al., 2013, Manikkam et al., 2012, Brieño-Enríquez et al., 2016, Titus et al., 2019), cocaine (Le et al., 2017, Vassoler et al., 2013), alcohol (Chastain and Sarkar, 2017, Legault et al., 2017) and malnutrition (Wei et al., 2014). Although only subtle changes in histone modifications in mammalian germline are currently observed related to epigenetic inheritance (Carone et al., 2010), a small portion of histone modifications are retained during the chromatin remodelling throughout spermatogenesis (Campos et al., 2014), suggesting to have a direct effect on gene expression in early stage embryos and therefore carriers of paternal information (Siklenka et al., 2015, Lesch et al., 2016). Recent studies also report that in mammalian male germline, diet-induced alteration of epigenetic information, in the form of tRNA-derived small RNAs (tsRNAs) and RNA modification, are directly linked to metabolic disorder, such as impaired

glucose tolerance and insulin resistance, in the offspring (Chen et al., 2016c, Zhang et al., 2016d). Interestingly, tsRNAs, or tRNA fragments (tRFs), as well as some microRNAs are found to be originally synthesized in the somatic cells of the epididymis, and later transferred by epididymosomes into sperm, representing a potential mechanism overcoming the physical barrier between somatic and germline cells (Sharma et al., 2017). This epididymosome-mediated transfer of sncRNAs, combined with degradation or dilution of piRNA after sperm entering epididymis, lead to another epigenetic reprogramming during spermatogenesis. The mechanism initiating this epigenetic reprogramming is largely unknown, but it is possible that environmental signalling may play a role in the regulation.

Proving that inheritable phenotypes have an epigenetic basis is very challenging (Blake and Watson, 2016). Given the fact that very few examples of heritable epigenetic variation at individual genes are found in mammals (Rakyan et al., 2002), it is more likely that subtle, but quantitative epigenetic changes, in different forms, will work in concert to respond to environmental cues experienced by former generations and lead to the altered metabolic or neurological phenotypes in offspring (Heard and Martienssen, 2014, Feeney et al., 2014, Murdoch et al., 2016, Triantaphyllopoulos et al., 2016). Also, alteration in epigenome, especially at telomere and repetitive elements, may lead to genomic instability which will cause inheritable phenotypic changes. To detect the potential carriers of epigenetic information, i.e. inheritable epigenetic changes in germline that links to environmental cues or observed phenotypes, and rule out naturally-occurring epivariations (Shea et al., 2015) and genetic causes , unbiased, comprehensive and costly studies are required to correlate epigenome, genome and transcriptome data of several generations. It is feasible but still difficult to obtain those data from limited amount of material such as individual preimplantation embryos or PGCs of individual embryos. Also, different epivariations may be established in each generation due to the changed epigenome in previous generation. No current evidence in mammals indicates a stable epivariation can be transmitted across generations, although in plants ‘epialleles’ have been discovered for more than half a century (Weigel and Colot, 2012) and a recent study also reports epialleles and the generation of ‘epiline’ in Drosophila (Ciabrelli et al., 2017).

In general, although epigenetic inheritance may play an important role in mammalian life cycles, providing direct evidence for its existence and uncovering the underlying mechanisms remains extremely difficult (Blake and Watson, 2016).

1.2.5.2 Pig as a model for studying heritable epigenetic response to malnutrition and obesity

Pig is a highly accessible and amendable large-animal model, which has multiple advantages over mouse or other large animals as an energy metabolism-related model, such as the proportional organ sizes, omnivorous habits, similar lipoprotein metabolism, and propensity to sedentary behaviour and obesity (Gonzalez-Bulnes et al., 2014). Indeed, paternal and maternal diet-induced transgenerational inheritance have already been reported in pig (Gonzalez-Bulnes et al., 2014, Braunschweig et al., 2012), although the underlying mechanism is unknown. Epigenetic mechanism is likely to contribute to the inheritable phenotypes, as pigs used in the studies have a balanced genetic background. Alternation in DNA methylation has also been observed in adult F2 in one study (Braunschweig et al., 2012). Epigenetic profiles of pPGCs from differentially treated ancestor may provide more information for understanding the transgenerational inheritance.

1.3 In Vitro systems for mammalian germline studies

The in vitro induction system of hPGCLCs circumvents the technical and ethical barriers of studying hPGCs for their specification and early development, and therefore provides valuable disease models and potential treatments for infertility and germ cell tumour (Chen et al., 2017a). In this section, methods of inducing PGCLCs from pluripotent stem cells (PSCs) without genetic manipulation (i.e. overexpression of core germline factors) in different mammalian species, as well as the applications of these systems will be introduced.

1.3.1 Generation and characterisation of PGCLCs in different species

In vitro induction of PGCLC from PSCs has been demonstrated in multiple mammalian species, including mouse, human, rhesus macaque (*Macaca mulatta*), *Cynomolgus* monkeys (*Macaca fascicularis*) and pig. As mentioned above, the signalling pathways inducing PGC specification are largely conserved between species. Same four cytokines, BMP4, LIF, SCF and EGF, are used in the most robust PGC differentiation media in mouse, which is further applied in human, non-human primate and pig. Although BMP4 alone can induce Blimp1-mVenus and Stella-ECFP positive ($BVSC^+$) PGC-like cells, LIF enhances the maintenance and proliferation of $BVSC^+$ cells. The combination of LIF, SCF and EGF further improve the robustness of PGCLC induction protocol (Hayashi et al., 2011, Ohinata et al., 2009). BMP8b is initially included in the media, which is suggested to work synergistically with BMP4. However, no significant difference is detected in PGC induction with or without BMP8b (Hayashi and Surani, 2009).

The main difference, and possibly the main effort endeavoured during inducing PGCLC in different species is to induce and capture the germline competent state, which is responsive to the BMP signalling, from different types of pluripotent stem cells. As shown in mice, pre-priming the naïve mESCs to mouse epiblast-like cells (mEpiLCs) with bFGF and Activin A, which exhibit the germline competent state, largely improve the mPGCLC from random differentiation to an induction efficiency of ~50% (Hayashi et al., 2011). Capture and characterisation of the germline-competent state is tricky. For example, while mEpiLCs which resemble the pre-gastrulating epiblast are germline-competent, mEpiSCs which show features of primed pluripotency and later primitive streak, exhibit fairly low conversion efficiencies towards mPGCLCs (Hayashi and Surani, 2009).

The situation in human was equally complicated, as different culture conditions may have generated a spectrum of pluripotent states. The first three protocols showing relatively high hPGCLC induction efficiency were introduced in 2015. Irie et al. converted hPSCs cultured in 4i condition, which was assumed naïve pluripotent cells, through either EpiLCs or directly as

aggregates in PGCLC media. They reported higher induction efficiency with direct differentiation from 4i hPSCs (Irie et al., 2015). Sugawa et al. started with hPSCs cultured in conventional MEF-conditioned media containing bFGF/KSR on matrigel-coated plates (i.e. feeder-free), which are considered primed/conventional pluripotent cells. These primed hPSCs were first induced to a primitive streak/mesodermal intermediate stage by a combination of Activin A, bFGF and also BMP4, and then form aggregates with BMP4 and LIF to induce hPGCLCs (Sugawa et al., 2015). Sasaki et al. induced primed hPSCs cultured in feeder-free and chemically-defined condition (StemFit meida) to incipient mesoderm-like cells (iMeLCs) with Activin A and glycogen synthase kinase3 (GSK-3) inhibitor (CHIR99021), and then to hPGCLC with the combination of BMP4, LIF, SCF and EGF (Sasaki et al., 2015). A similar protocol was also reported by Kobayashi et al., with primed hPSCs maintained in Essential 8 (E8) media and a different basal media (aRB27) used for the induction (Kobayashi et al., 2017a). The hPGCLC induction efficiency ranges from 10% to 50%, which varies largely with different hPSCs cell lines and induction protocols. Sasaki et al. further compared the transcriptional profiles of PGCLCs induced by different protocols and suggested that their BTAG (+) cells are similar to the TNAP/NANOS3(+) cells by Irie et al., while PGCLCs by Sugawa et al. exhibited a different expression pattern (Sasaki et al., 2015). They also argue that 4i hPSCs are in a similar state as iMeLCs, which is a peri-gastrulating epiblast-like state, instead of the assumed naïve state, as 4i hPSCs show upregulation of mesodermal markers but not consistent upregulation of naïve pluripotency genes. Notice that 4i media contains similar amount of WNT activator CHIR99021 with iMeLCs-inducing media. Treatment of hPSCs cultured in other conditions with CHIR99021 has also been demonstrated to efficiently induce mesendoderm differentiation marked by expression of primitive streak genes such as *T, MIXL1 and EOMES (Lam et al., 2014)*.

The Sasaki protocol was further modified to induce PGCLC from rhesus macaque iPSCs (riPSCs) (Sosa et al., 2018), with highest efficiency of 20% observed in D2 aggregates, while the Kobayashi protocol was modified to induce PGCLC in *Cynomolgus* monkeys (cyPGCLCs) (Kobayashi et al., 2017). The induction medium based on aRB27 was also used by Sakai et al. to

directly induce cyPGCLCs (i.e. without iMeLC induction) from PSCs, with the highest efficiency of 70% in D2 aggregates (Sakai et al., 2019). Surprisingly, it is shown that inhibition of WNT signalling by IWR1 in cyPSCs, which stabilizes their undifferentiated states, does not compromise the germline competence of these cells. This may reflect a distinct state of the cyPSCs used in the experiment (Sakai et al., 2019), as activation of WNT pathway is required for generating hPGCLC- and cyPGCLC-competent cells in previous study (Kobayashi et al., 2017). The two-step protocol in mice was modified to induce PGCLC from pig iPSCs (piPSCs) (Wang et al., 2016). The pig PGCLCs (pPGCLCs) exhibited germline features at both transcriptional and epigenetic level and was capable to further proliferate and differentiate in vitro and after xenotransplantation. However, pPGCLC showed high level of SOX2, which is not observed in in vivo pPGC, suggesting further validation of the cell identity is required.

1.3.2 **Applications of in vitro system of mammalian PGCLC**

1.3.2.1 Dissecting the mechanisms driving germline Specification

As mentioned in previous sections, regulatory factors for germline competency, such as EOMES, and PGC specification, such as SOX17, were identified using hPGCLC induction systems, which has revealed key differences compared to mPGC. A recent research reported a systematic strategy to investigate the germline ontogeny, of which a reporter system is used to track cell identities during mPGCLC induction and CRISPR screening is applied to identify essential regulatory genes for the cell fate transitions (Hackett et al., 2018). A single-copy of Cas9 was introduced and a library of exon-targeting guide RNA (gRNA) are used to transfect the reporter mESCs before PGCLC induction. After initiating PGCLC induction, cells with different identities are sorted based on the reporter signals. As gRNAs become depleted concomitant with cell loss, essential genes can be identified by comparing gRNA frequency to the preceding cell population. Regulatory genes for different events, such as acquisition of post-implantation epiblast fate and specification of PGC fate, were identified.

1.3.2.2 Disease model of infertility and germ cell tumour

The quantity and quality of germ cells have a significant impact on individual's fertility and also wellbeing of the next generations. Failure in PGC specification, or abnormality of germline development during embryonic stages was impossible to study before the establishment of hPGCLC induction system, and therefore hPGCLC provides model to better understand human fertility and germ cell tumours (Chen et al., 2017a). For example, a recent study reported the generation of hPGCLC from iPSC derived from primary ovarian insufficiency (POI) patient, which presents the possibility of exploring embryonic germline development in POI patient (Yang et al., 2019). Germ cell tumours are hypothesized to originate from lesions of arrested PGCs, which maintain pluripotency program and fail to further differentiate. Recent studies indicated that intracranial germinomas exhibited not only similar transcriptional program to hPGC, but also epigenetic features including global hypomethylation and enrichment of H3K27me3, suggesting they are originated from abnormal hPGC in Wk4 to Wk10 (Gell et al., 2018). Using in vitro induction system, Gell et al. also investigated the potential role of PRDM14 in abnormality of hPGC, as PRDM14 is highly expressed in germ cell tumours and embryonic carcinoma cell lines. Interestingly, constitutive overexpression of PRDM14 in PGCLCs does not affect PGC identity, but significantly increased the percent of EdU positive cells, i.e. proliferative PGCLCs (Gell et al., 2018). As later study suggests PRDM14 is required for hPGCLC induction, strict regulation of PRDM14 may be important for proper hPGC development.

1.3.2.3 In vitro gametes

Inducing PGCLC from PSCs is the first step of reconstituting the entire process of gametogenesis in vitro. Complete gametogenesis from individual-derived iPSCs has the potential to restore fertility to individuals who lost their germline due to injury or treatment for cancer. mPGCLC induced from PSCs were often tested for their capacity for gametogenesis. Capacity for spermatogenesis was tested via gonad reconstruction or seminiferous tubule injection (Hayashi et al., 2011), and functional spermatozoa were obtained and contributed to fertile offspring. The capacity for oogenesis has been tested via

generating mPGCLC-somatic cell aggregates in vitro (i.e. reconstituted ovaries), which is then transplanted under the ovarian bursa. The female PGCLCs after transplantation can develop into fully potent oocytes that contributed to healthy and fertile offspring (Hayashi et al., 2012). A more recent study further developed protocol for in vitro differentiation of reconstituted ovaries and generated fully potent mature oocytes from mouse PSCs completely in culture (Hikabe et al., 2016). Similar approaches were also applied for hPGCLC. For oogenesis, hPGCLC-mouse embryonic gonadal somatic cell aggregates were generated. However, hPGCLC survived but did not proliferate or differentiate further after transplantation in female mice (Sugawa et al., 2015). A more recent attempt showed that when aggregating with mouse embryonic ovarian somatic cells, hPGCLCs developed oogonia-like features at transcriptional and phenotypic level. However, those cells exhibited incomplete epigenetic reprogramming and failed to enter meiosis in vitro (Stringer and Western, 2019, Yamashiro et al., 2018). Furthermore, for spermatogenesis, non-human primate PGCLCs failed to fully reprogram and differentiate to spermatogonia in mouse or primate testicular niche, which was suggested due to either an incompetent primate PGCLC or primate PGCLC are incapable of differentiating in an adult niche. Overall, although the establishment of hPGCLC induction has created revolutionary opportunities in germ cell researches, in vitro gametogenesis in human remains a challenge.

1.4 Hypothesis and Aim

The hypothesis of this project is that epigenetic reprogramming in PGC and its transcriptional regulation are largely consistent between human and pig and therefore pig PGC is a suitable model for understanding human development. Since early hPGCs (Wk2-3) are not accessible, investigations of the epigenetic events during pig PGC specification can inform on fundamental mechanisms of germ cell development in humans and shed light into transgenerational epigenetic inheritance. The aim of this project is to profile new-specified and gonadal pPGCs at transcriptomic and epigenetic level with single cell RNA-seq, immunofluorescence and whole-genome bisulfite sequencing with post-bisulfite adapter tagging (PBAT-seq). This detailed profile of pPGC will also

be compared to the public data of mouse and primate PGCs to identify the conservation and divergence of mammalian PGC development.

Chapter 2 Transcriptional network of pre-/early migratory and gonadal pig PGCs

2.1 Introduction

As key differences have been reported between mice and hPGC development (Tang et al., 2015, Irie et al., 2015, von Meyenn et al., 2016) and pre- and early migratory hPGCs are not readily accessible, in vivo data of pre-/early migratory PGCs from other non-rodent animal models are required for a comprehensive understanding of mammalian PGC specification and development. Therefore, the single-cell transcriptome of 17 E14 (pre-/early migratory) and 22 E31 (gonadal) pPGCs, as well as 89 E14 neighbouring somatic cells and E11 epiblast (Ramos-Ibeas et al., 2019), were analysed in this chapter. The aim of this chapter is to examine the transcriptional network and epigenetic reprogramming in early and gonadal by analysing this high-resolution transcriptome data. Other PGC-featured events, such as cell cycle arrest, DNA repair, metabolic changes and unique membrane proteins, were also discussed. Those events contribute to proper specification, survival and development of PGC. To gain a better understanding of the similarities with other mammals, I performed comparative analysis with available datasets from human and Cy monkey.

2.2 Material and Method

2.2.1 Embryo and Gonadal tissue collection

All the procedures involving animals have been approved by the School of Biosciences Ethics Review Committee, The University of Nottingham. Embryos were retrieved from crossbred Large White and Landrace sows (2–3 years old) between days 11 to 35 after artificial insemination. Later stage embryos (>E35) were retrieved from abattoir and the stage of the embryo was estimated by cranial‐caudal length. E11 and E14 embryos were flushed from the uterine horns with warm washing buffer (PBS supplemented with 1% fetal bovine serum (FBS)). Later stage embryos (>= E25) were manually dissected

from the uterine horns and washed with washing buffer. Epiblast from E11 embryos were manually dissected and stored at -80°C before further processing for LC-MS (see below).

2.2.2 Immunofluorescence

2.2.2.1 Fixation and cryostat of embryo and gonadal tissue

Embryos were processed as previously described (Kobayashi et al., 2017). Briefly, embryos and gonads were fixed in 4% paraformaldehyde (PFA)/PBS overnight (ON) at 4°C. Fixed embryos were incubated in 30% sucrose/PBS for two days at 4°C prior to mounting in optimal cutting temperature (OCT) compound. Cryosections were cut at 5-7 µm onto Superfrost plus glass slides. Sections were left to air dry for 1-2 h before IF or storage in -80°C.

2.2.2.2 Staining

For IF, air-dried cryosections were washed with PBS for 10 mins to remove OCT compound. Antigen retrieval was then performed by boiling the slides in 0.01M Citrate Buffer (pH 6.0) for 10 min. Sections were permeabilized with 1% Triton-X100 in PBS for 15 min. Triton-X100 was washed three times for 5 min each, and blocking solution (PBS supplemented with 5% BSA and 10% Donkey serum) was added for 1.5 h. After blocking, sections were incubated with the desired primary antibody (Details about primary antibodies and their working dilution please see Appendix Table 1) ON at 4°C in a humidified chamber. Slides were then washed three times with 0.1% Tween-20/PBS. Slides were then incubated with fluorescent (Alexa Fluorophore 488, 555, and/or 647; Invitrogen)-conjugated secondary antibodies for 40 min at room temperature (RT). Slides were mounted with Fluoroshield with DAPI (Sigma) and sealed with nail varnish. Slides were kept at -20 °C until observed.

2.2.2.3 Imaging and post-imaging processing

Image acquisition was performed using SimplePCI capture software on an epifluorescence microscope (Leica). Fiji was used for cell count and

fluorescence quantification of ROI (Schindelin et al., 2012). For fluorescence quantification, background intensity was subtracted to generate corrected total cell fluorescence (CTCF), i.e. CTCF = Integrated Density – (Area of selected cell X Mean fluorescence of background readings) (McCloy et al., 2014).

2.2.3 Tissue dissociation and Fluorescence-activated cell sorting (FACS) for PGC collection

Pig PGC isolation was carried out as previously described (Hyldig et al., 2011a). Briefly, embryos between E14 to E35 were stored in DMEM/F-12 supplemented with 40% FBS at 4°C overnight before being processed the next day. Dissected posterior ends of E14 embryos containing PGC clusters and gonads from E31 and E35 embryos were digested at 37 °C for 30 mins using Collagenase IV (2mg/ml in DMEM), with gentle pipetting every 5 mins. The cell suspension was washed with DMEM, centrifuged and the pellet re-suspended in TrypLE Express (GIBCO) for further digestion at 37°C for 3-5 mins. Enzymatic digestion was neutralized with dissection medium (DMEM/F-12 with 10% FBS, 25 mM HEPES and 100 U/ml Penicillin-0.1 mg/ml Streptomycin). The cell suspension was filtered through a 40 µm cell strainer into FACS tube. Following centrifugation, cells were re-suspended and incubated in dissection medium with Sda/GM2 antibody (Klisch et al., 2011) for 30 mins at 4°C. After washing with dissection medium, cells were re-suspended and incubated in dissection medium with Alexa-488 Donkey Anti Mouse (Invitrogen) for 30 mins, and then diluted with dissection media and FACS sorted by MoFlo XDP Cell Sorter (Beckman Coulter).

2.2.4 Single-cell library preparation

FACS sorted cells were washed in a small drop of PBS-PVP and single cells were manually collected with thin capillaries and placed into PCR tubes to prepare single-cell cDNA libraries following the Smart-seq2 protocol (Picelli et al., 2014).

Single cells were lysed by incubation at 72 °C for 3 min in PCR tubes containing 4 µl of cell lysis buffer, oligo-dT primer and dNTP mix. Reverse

transcription and PCR pre-amplification were carried out with SuperScript II (Invitrogen) and KAPA HiFi HotStart ReadyMix (KAPA Biosystems) respectively according to Picelli et al. protocol (Picelli et al., 2014). PCR products were purified using Ampure XP beads (Beckman Coulter), and library size distribution was checked on Agilent dsDNA High Sensitivity DNA chips on an Agilent 2100 Bioanalyzer (Agilent Technologies). Concentration was quantified using Qubit Quant-iT dsDNA High-Sensitivity Assay Kit (Invitrogen). Samples with more than 0.2 ng μl^{-1}, free of short fragments (<500 bp) and with a peak at around 1.5–2 kb were selected for library preparation with Nextera XT DNA Library Preparation Kit (Illumina). Tagmentation reaction and further PCR amplification for 12 cycles were carried out, and PCR products were again purified using Ampure XP beads. Quality of the final cDNA library was analysed on an Agilent high sensitivity DNA chip. Final cDNA libraries had an average size of 700–800 bp and were quantified using NEBNext Library Quant Kit for Illumina (New England BioLabs) following the manufacturer instructions. Finally, libraries were pooled in groups of 50 with a 2 nM final concentration, and DNA sequencing was performed on a HiSeq 2500 Sequencing System (Illumina).

2.2.5 **Bioinformatic analysis**

2.2.5.1 Processing of raw data

Raw PE reads were trimmed against adaptor sequences by scythe (v0.981), and quality-trimmed by sickle (v1.33) using default settings. Trimmed reads were directionally aligned to the pig genome (Sus scrofa v11) by hisat2 (v2.1.0) with *-know-splicestie-infile* setting to increase mapping accuracy of splicing reads. Uniquely and correctly mapped reads were extracted for the downstream analysis. htseq-count was used to count the number of reads aligned to each gene (Sus scrofa v11.2 Ensembl annotation build 91). Gene expression level was calculated and normalised by Transcripts Per Kilobase Million (TPM).

Low quality cells were filtered out from the dataset to reduce the downstream analysis noise. First, the total number of reads mapped to gene transcripts was calculated for each cell, and those with less than 1 million were removed. Second, the proportion of reads aligned to mitochondrial genes was estimated,

as a high proportion suggests poor quality cells (Ilicic et al., 2016). The proportion cut-off was set at 0.5. Only cells of proportions below 0.5 were kept for the next analysis. Third, 2 outlier cells were identified by t-SNE dimensionality reduction. A total of 14,873 out of 25,880 annotated genes were identified in at least 3 cells with TPM > 1.

2.2.5.2 Biological analysis

2.2.5.2.1 Clustering

The R package "scater" was applied to normalise read counts of genes for each good quality cell with acceptable sequencing coverage. A non-linear approach, t-stochastic neighbour embedding (t-SNE), was used to identify the relations between cells using normalised read counts. Unsupervised hierarchical clustering using all expressed genes as input was conducted on all filtered cells by normalised read counts in log2 scale. The distance method was euclidean, and the cluster method was ward.D2.

2.2.5.2.2 Differential expression and enrichment analysis

Pairwise comparisons of single-cell differential expressions were performed by SCDE using normalised read counts among four embryo stages. Two-tailed adjusted p-value were calculated using cZ scores from Benjamini–Hochberg multiple testing corrections, which followed a normal distribution. Significantly expressed genes were selected with a p-value <0.05 as the threshold. Euclidean distance and default hclust were applied to determine the relationships between cells and between genes. Gene Ontology (GO) gene set enrichment analysis with DEGs utilised goseq for each pairwise comparison, also with upregulated DEGs and downregulated DEGs separately. GO term annotation was retrieved from the Ensembl database (Sus scrofa v11.1 Ensembl annotation version 91). Enrichment analysis of biological pathways (KEGG) was performed with DEGs by R package "clusterProfiler". Ensembl gene IDs of DEGs were mapped to NCBI gene IDs for KEGG pathway prior to enrichment analysis.

2.2.5.2.3 Inference of embryonic sex

Expressions of all the single-copy genes on chrY were summed up to determine the gender of each cell. First, any cell with the total TPM of chrY single-copy genes ≥ 10 was regarded a male cell. Others were regarded as female cells. Then, the ratios of the total gene expressions between chrY and chrX (∑ ChrY Total TPM / ∑ ChrX Total TPM) were calculated across all cells. Any pre-determined male cell with the ratio lower than the maximal ratio of pre-determined female cells was regarded as the female cell.

2.2.5.2.4 Chromosome X dosage compensation analysis

Genes of chromosome X and three autosomes (chr1, chr2, chr3) were extracted, and the geometric mean TPM of chromosomal expressed genes was calculated for each cell separately. Then the overall geometric mean TPM was obtained for each developmental stage by embryo sex, as well as the total TPM. Each TPM value was incremental by one (TPM + 1) for the calculation of geometric mean TPM. Only shared expressed genes between female and male cells were taken into account in the calculation of female/male expression ratio for each chromosome. Median Female/Male expression ratio was estimated for each stage across the whole X chromosome with 1 Mb window. The ratio of chrX/auto in each cell was inferred by the median value of bootstrapped ratios. Each ratio was estimated by the total TPMs of a certain number of random-selected genes. The median ratios were grouped by embryo sex.

2.2.5.2.5 Analyses of allelic expression

Trimmed reads were aligned to chromosome X of the pig genome (Sus Scrofa v11.1) by hisat2. Duplicated reads were marked by picard (v2.12.1). GATK (v3.8) was used to retrieve allelic read counts for SNVs annotated in dbSNP. Only validated SNVs (dbSNP flag VLD) were extracted for downstream analysis. SnpEff (v4.3) was applied to annotate called SNVs with Sus scrofa v11.1 Ensembl annotation. Low coverage SNVs (<3 reads) were excluded from the analysis, and we only kept SNVs that occurred at least in two different cells for each stage. The expressions of mono-/bi-allelic genes were inferred based on SNVs in each female cell of each stage.

2.2.5.2.6 Single cell trajectory analysis

Trajectory modelling and pseudotemporal ordering of cells was performed using TPM data with Monocle 2 (Qiu et al., 2017) (version 2.12.0). Top 1000 significant differentially expressed genes between clusters were used for ordering the cells.

2.2.5.2.7 Comparison of pig, human and cynomolgus monkey datasets

In total, dataset of E14-31 pig cells (128 from our study), processed data of Wk4-7 human cells (149) retrieved from GSE86146 (Dolatabadi et al., 2017) and processed data of E13-55 cy monkey cells (100) retrieved from GEO: GSE76267, GSE74767 and GSE67259 (Sasaki et al., 2016) were included in the comparison. Natural log-transformed relative expression (i.e. ln(TPMs+1) in pig and human, ln(RPMs+1) in cy monkey) of common genes (i.e. homologues genes with same gene name) across three species were imported and processed by *FindIntegrationAnchors* and *IntegrateData* (k.filters set as "NA") functions in Seurat (version 3.1.2)(Stuart et al., 2019). Dimensionality reduction by UMAP was then performed for the integrated dataset.
Expression of selected lineage markers and epigenetic modifiers in E14-31 pig cells, Wk4-7 human cells and E13-55 *Cynomolgus* cells were plotted separately with pheatmap package.

2.2.5.2.8 Cell cycle analysis

Default settings of *CellCycleScoring* function in Seurat were used to score the cell cycle phases of each single cell. In brief, single cells were assigned a score with *AddModuleScore* function based on its expression of G2/M- and S-phase markers provided in Seurat. The single cells highly expressing G2/M- or S-phase markers were assigned as G2/M- or S-phase cells, respectively, and the single cells not expressing any of the two categories of genes were assigned as G1 phase.

2.2.5.2.9 Signature set analysis

With the processed single cell RNA-seq data of pig embryos from Ramos-Ibeas et al. (Ramos-Ibeas et al., 2019) we used *FindMarkers* function in Seurat (Wilcoxon rank sum test) to identify the highly expressed genes (avg_logFC >=1 and adjusted.p_val <=0.05) as the signature set in E6 ICM, E8 epiblast and E11 epiblast. Next, we calculated the relative average expression level of each signature set with *AddModuleScore* function of Seurat in single cells of E14 Soma, E14 PGC and E31 PGC, which was then visualized by heatmap using pheatmap package.

2.3 Results

2.3.1 **Single-cell transcriptome of pre-migratory and gonadal pPGC reveal highly dynamic states of PGC development**

2.3.1.1 Isolation of pig PGC with Sda/GM2 marker from E14 embryos and embryonic gonad

Prior to single cell library preparation, pPGCs were isolated from E14 embryos and E31 gonads using FACS with a home-made antibody targeting Sda/GM2 epitope, a specific pig PGC marker (Klisch et al., 2011). The antibody specificity and sensitivity for pPGCs were confirmed by immunofluorescence (IF) staining of tissue sections (Figure 2-1). Sda/GM2 staining was shown overlapped with PGC marker SOX17 and BLIMP1 as well as pluripotency marker NANOG at the posterior end of 2-4 somite stage embryo (E14) and overlapped with SOX17 and OCT4 at early gonadal stage (E31). Secondary antibody control (i.e. Secondary-only) is used during FACS to determine the sorting gate. Sda/GM2-positive and -negative cells were collected separately after FACS and manually isolated into single cells using mouth pipette, which are then processed following the Smart-seq2 protocol (Picelli et al., 2014).

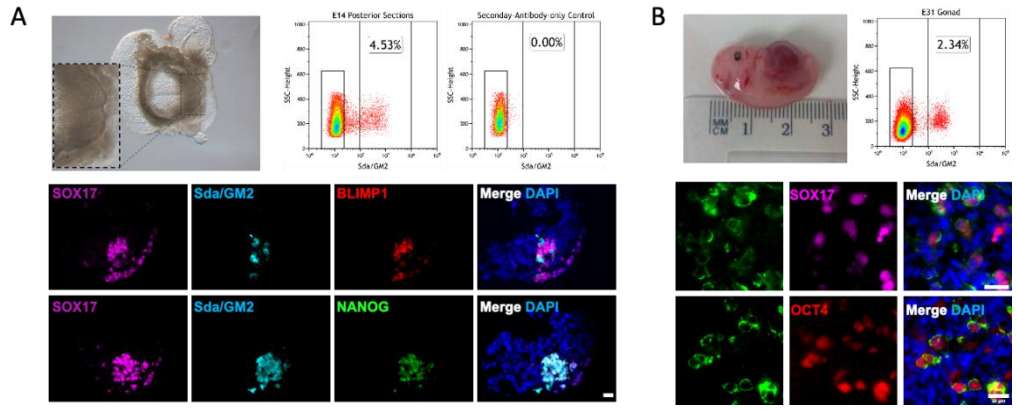


Figure 2-1 Confirmation of the specificity and sensitive of Sda/GM2 towards pPGC in pre-migratory (A) and gonadal stages (B) via FACS (top) and IF (bottom). Sda/GM2^{+} pPGCs are co-stained with known PGC markers (A) SOX17, BLIMP1 and (B) OCT4. Scale: 20um.

2.3.1.2 Confirmation of PGC identity and stage

2.3.1.2.1 Cells with different identities show clear separation with each other

As mentioned in introduction, the transcriptome dataset contains cells collected from E11 (Ramos-Ibeas et al., 2019), E14 and E31 embryo. Low-quality cells and genes that are not detected were first removed from the transcriptome. In total, 175 single cells and 13251 genes were used for the downstream analysis. With the mapped and filtered transcriptome, UHC and t-SNE analysis were performed to show segregation and clustering of different cell types. Clear separation was determined between PGCs (17 of E14 PGC and 22 of E31 PGC), E14 somatic cells (n=89) and Sph Epi (n=47), while two PGC populations (E14 PGC and E31 PGC) cluster closely with each other (Figure 2-2 and Figure 2-3). No separation due to the gender of cells was observed in any of those cell types. The expression profile of known pluripotency and lineage markers further validate the identity of cells (Figure 2-2). E11 Epi cells show high level of pluripotency markers, and low or absent expression of germline or somatic-lineage markers. Known PGC markers, including *PRDM1*, *TFAP2C*, *NANOS3* and *KIT,* as well as pluripotency markers *NANOG* and *POU5F1*, are highly expressed in both E14 and E31 PGC clusters, whereas

late PGC markers such as *DAZL*, *DDX4* and *PIWIL2* are only detected in E31 gonadal PGCs. The expression of PRDM14 is only detected in a few pPGCs, mainly at gonadal stage, which is consistent with previous IF detection of PRDM14 only in gonadal pPGCs (Kobayashi et al., 2017) and previous bulk RNA-seq data (Figure 2-2). As expected, E14 somatic cells show high level of somatic lineage and meso/endodermal genes, such as *HOXA1*, *HAND1*, EOMES and *GATA2*. Note that E14 somatic cells branch into two clusters, E14.soma.g1 and E14.soma.g2. Only E14.soma.g1 show expression of primitive streak gene T and EOMES and endodermal marker CXCR4, GATA4 and GATA6. Mesodermal marker *HAND1, GATA2* and *KDR* are expressed in both clusters. The expression pattern of those somatic lineage markers suggest E14.soma.g2 cluster may be more advanced than E14.soma.g1, while E14.soma.g1 cluster still possesses primitive streak features.

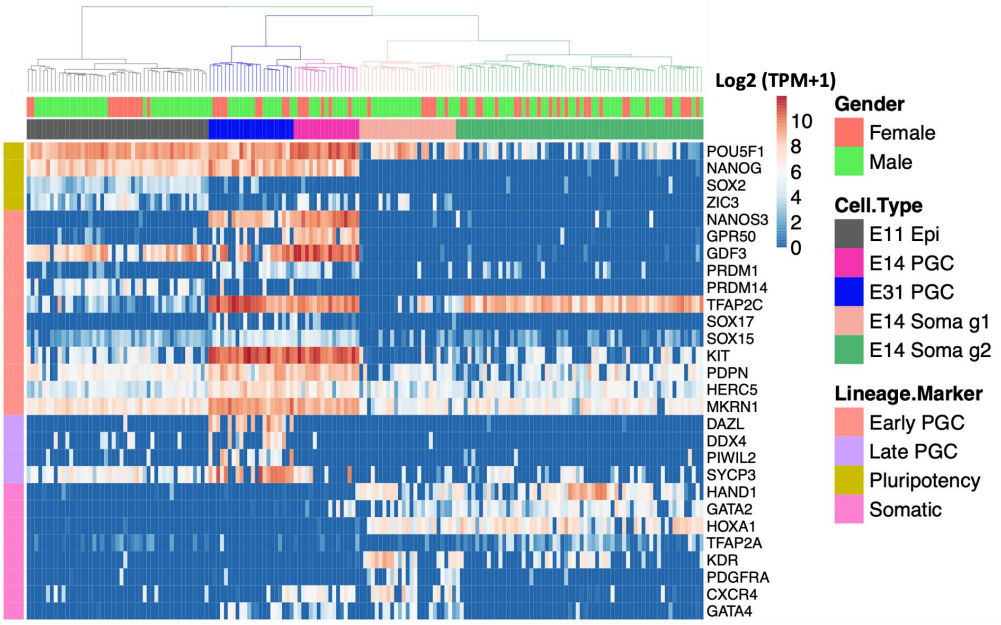


Figure 2-2 Expression heatmap of lineage markers in all samples. pPGCs express high level of known human PGC markers OCT4 (POU5F1), NANOG, BLIMP1, NANOS3, KIT, and SOX17, but majority lack expression of SOX2.Note that the UHC dendrogram is plotted based on all expressed genes.

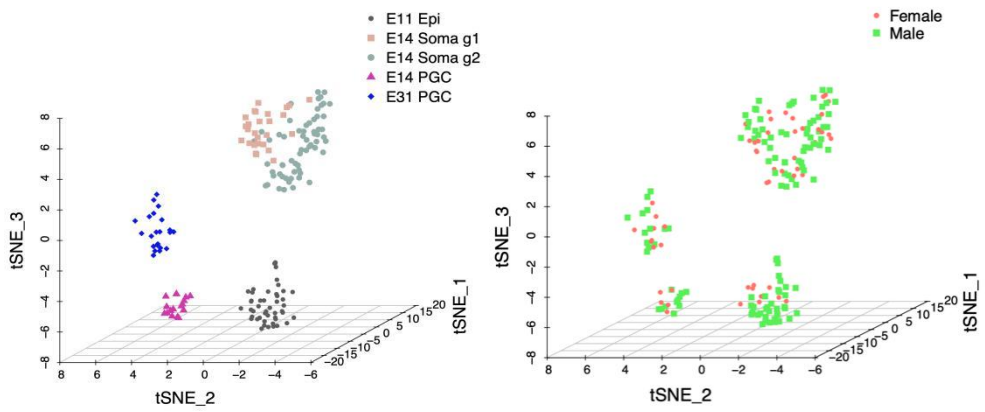


Figure 2-3 tSNE plot of all samples showing clear separation of E11 Epiblast, E14 Somatic cells and PGCs (left) but no clustering based on the gender of cells (i.e. Female and male cells within each identity mixed together) (right).

2.3.1.2.2 Genomic alignment of SOX17

Surprisingly, mapping of SOX17 to Ensembl reference in 2018 only detected expression in eight out of 39 pPGCs, which is inconsistent with the high level of SOX17 previously detected in majority of pPGC via IF (Kobayashi et al., 2017). The expression of SOX17 detected by bulk RNA-seq also does not show large difference between pPGC and E14 somatic tissues either, validating the observation by single cell RNA-seq. To examine if the lack of SOX17 in pPGC is due to the inconsistency between the SOX17 reference and SOX17 transcripts in pPGC (i.e. potential alternative splicing), alignment to extended SOX17 reference and de novo sequence assembly were performed. As SOX17 was detected in pig hypoblast but not epiblast by IF (Ramos-Ibeas et al., 2019), same analyses were performed for pre-gastrulation pig embryos (E8 and E11 Epiblast and hypoblast) to examine if same problem also occur in other cell types. As expected, using extended reference revealed more pPGCs and pig hypoblast cells with SOX17 expression, while only one extra cell is detected in E14 somatic cells, E8 and E11 epiblast (i.e. SOX17-negative samples) (Figure 2-4). This suggests that alignment with extended reference retrieved SOX17-expressing cells that were not detected by alignment with the

2018 reference but did not increase noise for SOX17 alignment. De novo assembly using reads from all SOX17-expressing cells (PGCs and hypoblast of all stages) generated two versions of assemblies (SOX17 de novo assembly v1 and v2), which both show extra transcript sequence on the 3' end compared to SOX17 cDNA reference or genomic sequence. De novo assembly using reads from hypoblast cells (SOX17 de novo assembly (E8-E11 cells)) and PGCs (SOX17 de novo assembly (E14-E31 cells)) generated shorter assemblies compared to SOX17 de novo assembly v1 and v2, which is presumably due to less reads used in the analysis. All assemblies generated only partially overlapped with exon2 of SOX17 and did not overlap with exon1. As qPCR primers spanning exon1 and exon2 showed amplification of cDNA from hypoblast (previous data in the lab), we assumed this was due to 3' -prime bias of Smart-seq2 protocol (Baran-Gale et al., 2017).

We did not further investigate the extra transcript sequence at the 3'end of the de novo assembly, as the most recent Ensembl Sus scrofa Genebuild, which was last updated/patched in Jun 2019, seems to largely extend the SOX17 reference sequence. The current SOX17 reference covers our de novo assemblies, which further validates the re-alignment and de novo assembly results (Figure 2-5).

Due to limitation of time, the re-alignment was not performed for the entire transcriptome in this project. However, the raw dataset of this scRNA-seq experiment has been made accessible to the public (GSE155136) and can be re-analysed when new genome assembly and/or alignment methods are released.



Figure 2-4 Comparison of SOX17 expression in the listed samples when different reference sequences were used. 'SOX17_ref', alignment using the Ensembl reference sequence in 2018; 'SOX17_extend', alignment using the Ensembl reference sequence in 2018 plus 5kb extension at both end. More reads are detected with the "Sox_17" reference.

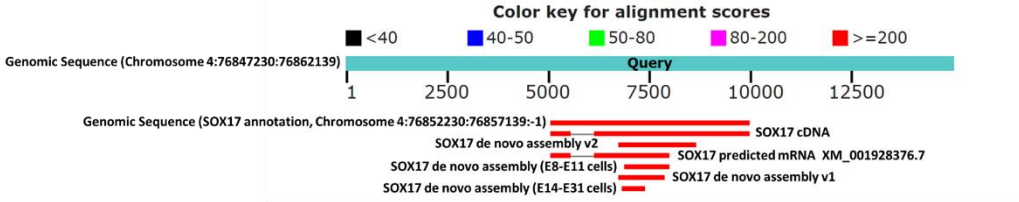


Figure 2-5 Comparison between the current Sox17 genomic sequence, cDNA, predicted mRNA and our assemblies. The current SOX17 genomic sequence covers the de novo assembly of SOX17 from transcriptome.

2.3.1.2.3 Trajectory inference and the identities of E14.soma.g1 and g2

We further explored the developmental processes in epiblast cells, PGC and somatic cells by trajectory and pesudotime inference. Using Monocle package with E11 Epiblast cells set as start point, the inferred developmental trajectory and pseudotime ordering shows that E14.soma.g1 is the more similar lineage to E11 Epiblast at transcriptomic level and E11 Epiblast progressed into E14.soma.g1, which further branches into E14.soma.g2 and PGC lineage (Figure 2-6). These suggest the E14.soma.g1 cells can be late primitive streak or early mesendodermal cells. The transition from E14.soma.g1 to E14.soma.g2 appears to be continuous, while a gap occurs around the branching point for PGC lineage. This suggests that transition state(s) may exist between the late primitive streak or early mesendodermal state we captured (i.e. E14.soma.g1) and the early-specified PGC.

The identity of E14.soma.g1 as late primitive streak or early mesendodermal cells was further supported by expression of transcriptional factors *T* and *EOMES* and *MESP1* (Soibam et al., 2015) and the co-expression of surface marker *CDH2* (NCAD), *CXCR4* and *PDGFRA* (Kopper and Benvenisty, 2012)(Figure 2-7). DEG analysis between E14.soma.g1 and g2 further reveals higher expression of pluripotency factor *OCT4* (Pan et al., 2016, Wen et al., 2017), WNT singaling components WNT5A, WNT5B, LEF1 and FZD10 (Kemp et al., 2007, Wu et al., 2015, Soibam et al., 2015), and downregulation

of mesendoderm impeding factor BACH1 (Wei et al., 2019) in E14.soma.g1. The upregulation of EMT driver SNAI1 and ZEB2, as well as downregulation of E-caderin (CDH1) (Pan et al., 2016) further suggest that E14.soma.g1 may have activated the Epithelial–mesenchymal transition (EMT) programme (Stryjewska et al., 2017, Chng et al., 2010, Stemmler et al., 2019). GO term enrichment analysis reveals that E14.soma.g1 is featured with "gastrulation" and "somite development", which is consistent with their late primitive streak identity, while E14.soma.g2 show features such as "cell migration" (Figure 2-8). Surprisingly, E14.soma.g2 also shows ectodermal feature, such as "epidermal cell differentiation" and "neuron projection morphogenesis". However, most of the commonly-used neuroectoderm markers (e.g. *PAX3, PAX6, SOX1, NES*, *NOG, FGF5,OTX2, NEUROG3, NKX6-3, SIX3, HESX1* and *CHRD)* are not present or expressed at low level in E14.soma.g2, suggesting they are unlikely to be neuroectoderm. On the other hand, E14.soma.g2 show upregulation of epidermal lineage markers (e.g. *KRT7,KRT8, KRT18, OVOL1, GATA3, TFAP2C* and *TP63*) (Li et al., 2013a, Liu et al., 2018, Nair et al., 2006, Soares and Zhou, 2018, Saykali et al., 2019, Krendl et al., 2017, Tchieu et al., 2017). Most of those markers are also present or upregulated in extra-embryonic lineage (Li et al., 2013b). A recent study has studied multiple lineages in in vitro cultured Cy monkey embryos, including the amnion lineage. Using amnion marker genes reported by this study as an signature set, I examined the relative expression of amnion signature set (shown as module score) in E11-31 pig cells. A large portion of E14.soma.g2 exhibits evidently higher expression of amnion markers compared to other cells, suggesting these cells may be amniotic precusors (Figure 2-9).

This analysis shows that there is a close spatial relationship between pre-migratory pPGCs, mesoderm and amnion precursors. Previous studies have shown that after their induction in the posterior pre-PS epiblast the PGC cluster localizes at the embryonic and extraembryonic border in pig pre-somitic embryos (Wolf et al., 2011, Kobayashi et al., 2017). These cell types are all induced by BMP signalling, which is high in the posterior end of the pig embryo (Yoshida et al., 2016, Valdez Magaña et al., 2014).

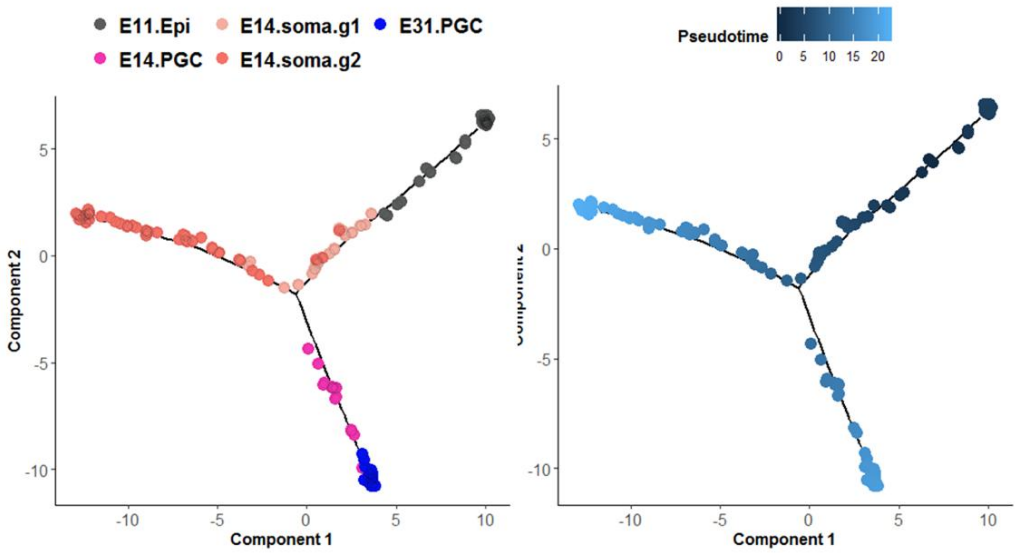


Figure 2-6 (Left) Trajectory of E11 epiblast, E14 cells and E31 PGC. (Right) Pesudotime of E11 epiblast, E14 cells and E31 PGC. E11 Epiblast progressed into E14.soma.g1, which further branches into E14.soma.g2 and PGC lineage.

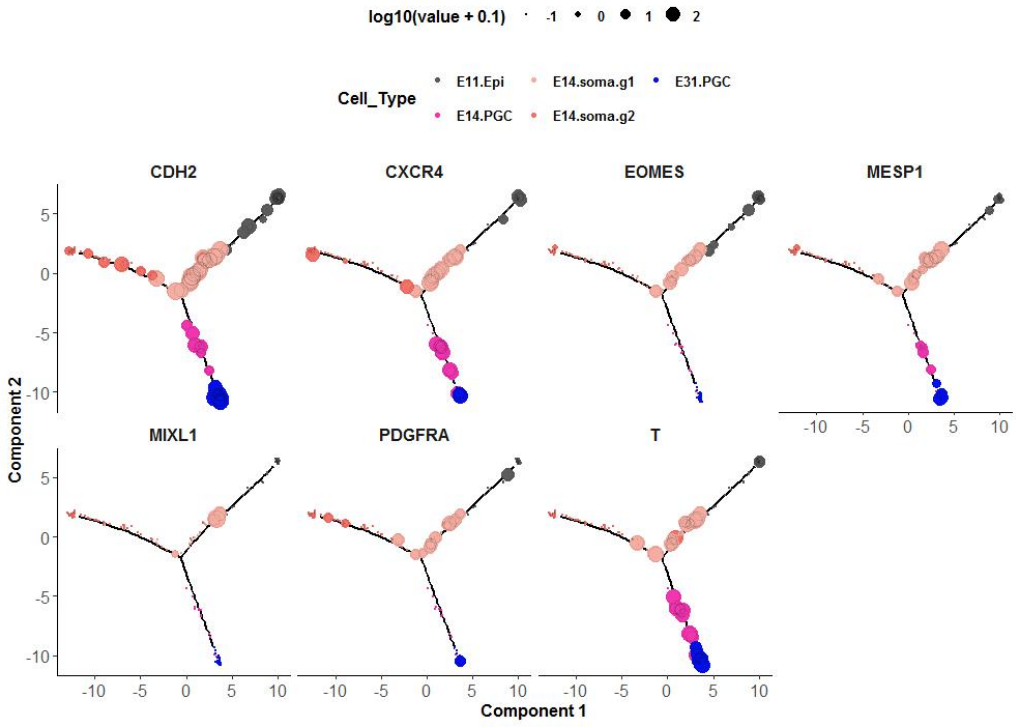


Figure 2-7 Expression of primitive streak markers EOMES, MIXL1 and T (Brachyury)), as well as mesendoderm surface marker CDH2 (NCAD), endodermal surface marker CXCR4, and mesodermal surface marker PDGFRA on trajectory of all samples. High expression of those genes in

E14.soma.g1 suggest it may be late primitive streak and/or early mesendodermal cells.

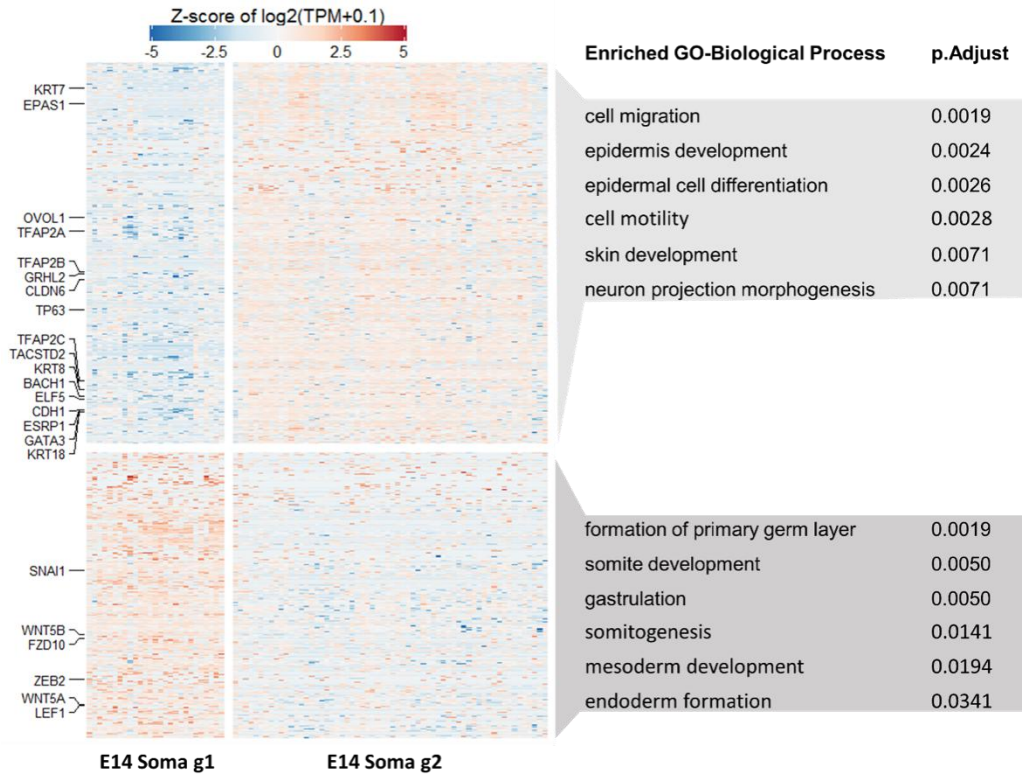


Figure 2-8 Expressional heatmap of DEGs between E14.soma.g1 (left) and g2 (right). A few signalling components and epidermal/extra-embryonic lineage markers are labelled on the left side. E14.soma.g1 and E14.soma.g2 are enriched for distinct GO terms, suggesting their very different identities.

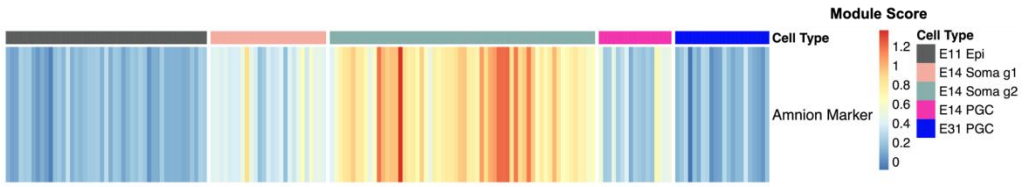


Figure 2-9 Relative average expression (module score) of amnion markers in E11-31 pig cells. Some E14.soma.g2 cells showing enrichment of amnion markers suggests they may be amnion precursors.

2.3.1.2.4 Determination of the cell cycle phase in single PGCs

A previous study suggested that pre-migratory pPGC showed no Edu incorporation, i.e. stopped DNA synthesis, shortly after specification (Kobayashi et al., 2017). It is possible that the E14 pPGCs in our dataset are at that specific stage. Therefore, attempts were made to determine the cell cycle phase in single PGCs cells of the dataset.

First, cell-cycle status was analysed using the expression profiles of previously reported G1/S and G2/M phase-specific genes (42 and 54 respectively) (Tirosh et al., 2016). Clustering of G1/S- and G2/M-specific genes showed separation according to their cell cycle-specificity (Figure 2-10). However, majority of E14 pPGCs did not exhibit feature of cell cycle arrest, i.e. no or low expression of both G1/S and G2/M-specific gene, and in general did not show an expression pattern different to E14 somatic cells.

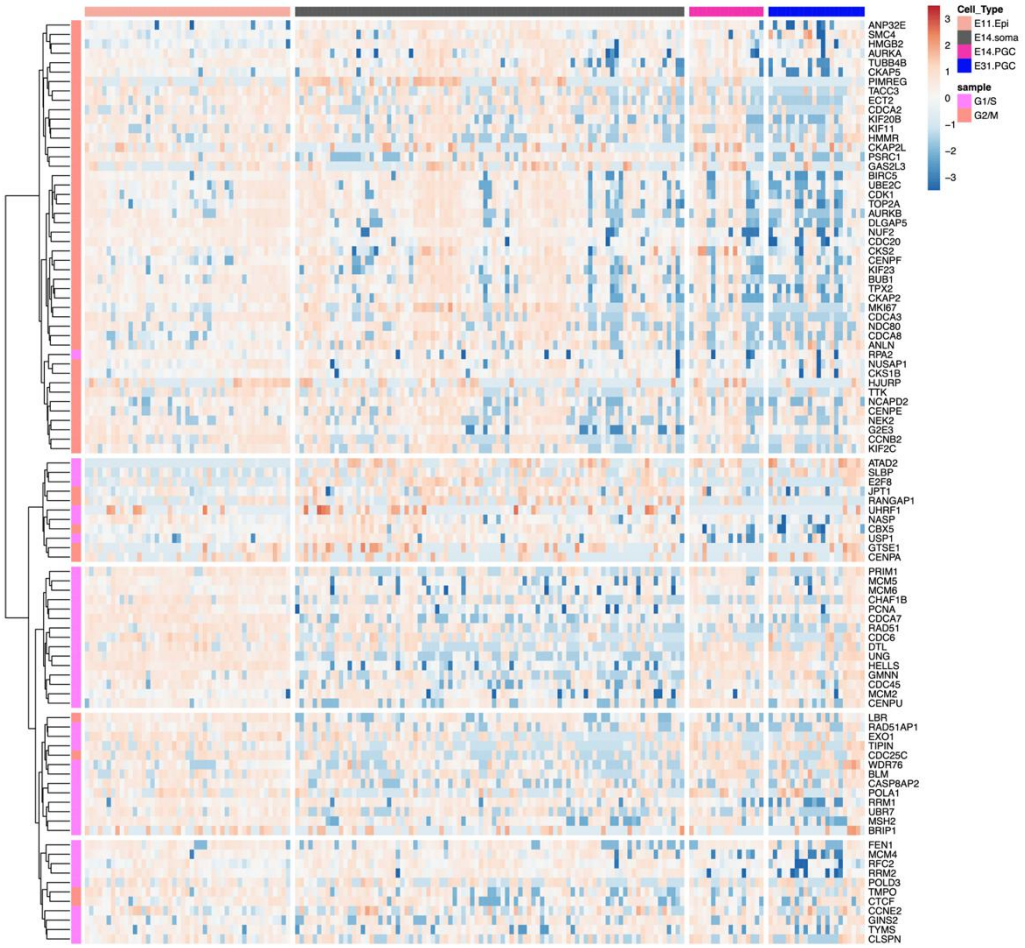


Figure 2-10 Heatmap showing expression of cell cycle related genes reported by Tirosh et al. in all samples. The unit of colour scale is Z score of log2 (TPM+1). No clear difference is observed between PGC and somatic lineages.

With these same G1/S and G2/M phase-specific gene sets, I also used *CellCycleScoring* function in *Seurat* package (Butler et al., 2018) to assign cells to discrete cell cycle states and performed PCA to test if cells would be separated by cell cycle status. Based on the *CellCycleScoring* results, E14 PGCs contained nearly 30% of S phase cells. Also, cells with different identities did not separate in the PCA plot, suggesting cells cannot be separated based on the previously reported G1/S and G2/M-specific genes (Figure 2-11).

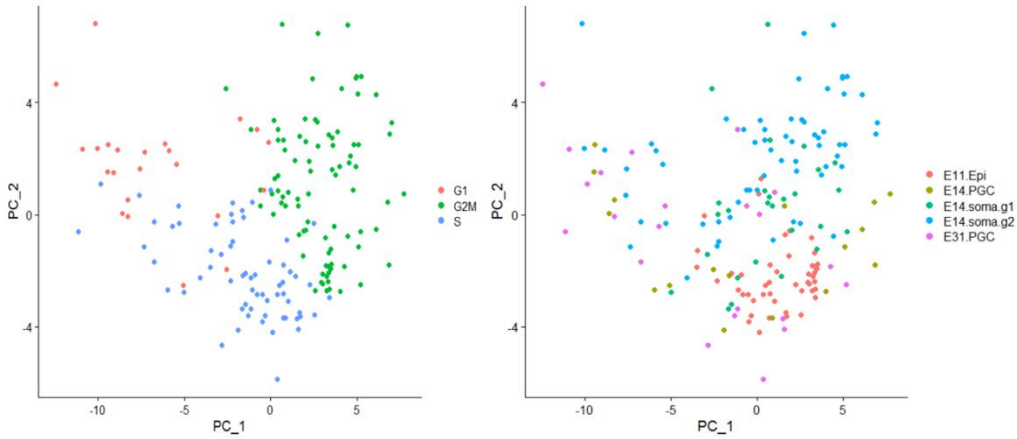


Figure 2-11 PCA based on the cell cycle genes reported by Tirosh et al. Each dot represents one single cell, which is coloured by cell cycle states (left) or by cell type (right)

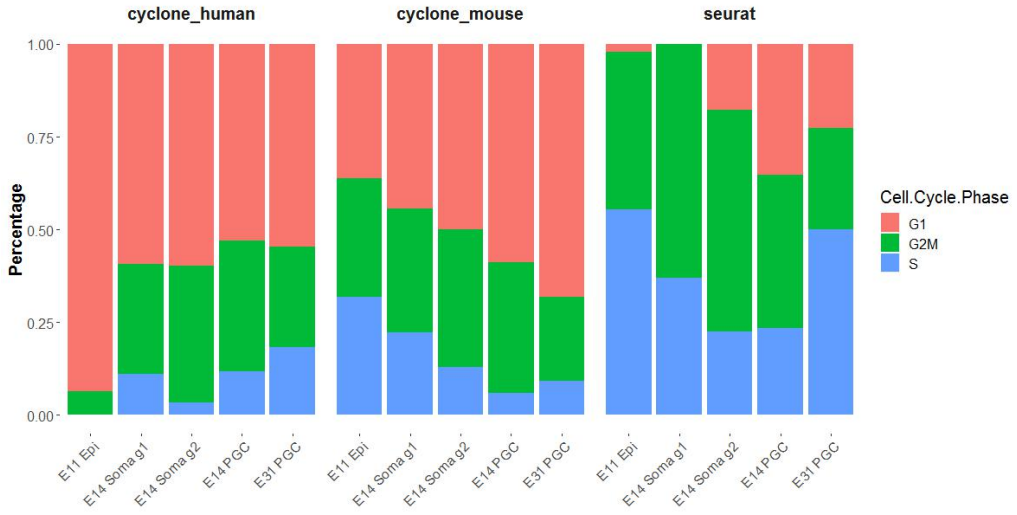


Figure 2-12 The assignment of cell cycle states by cyclone using human (left), using mouse (middle) and by Seurat (right) show clear difference in proportions of G1, S and G2M. Cyclone appears to preferably assign cells to G1 phase, while Seurat to S phase.

Another method, *Cyclone* classifier in *Scran* package (Scialdone et al., 2015) was also used to classify single cells into their cell cycle phases. As no scRNA-seq dataset in pig with experimental cell-cycle annotation is available for training the classifier, pig orthologues of pre-trained set of marker pairs for

mouse (654 orthologues) and human data (1239 orthologues) were used. However, the classified cell cycle phase for each cell varies largely between using mouse or human pre-trained sets. The difference is even starker when compared to the *CellCycleScoring* (Seurat) results (Figure 2-12). Compared to results from Cyclone, more cells are assigned to S and G2 phase and less assigned to G1 phase in the Seurat results.

The bioinformatics tools tested above failed to generate consistent results for cell cycle classification. *Cyclone* classifier and *CellCycleScoring* function are currently the only two bioinformatic methods assigning single cells to discrete cell cycle states. However, a recent study reported that neither methods agreed well with the cell cycle phase measurement by fluorescent ubiquitination cell cycle indicators (FUCCI) (Hsiao et al., 2020). These suggest that further research and better training sets, such as pig datasets with experimental cell-cycle annotation, are required to accurately classify the cell cycle phase based on transcriptional data.

2.3.1.3 Identification of Differentially Expressed Gene in PGC compared to Somatic cells and Epiblast cells

In the following sections, I focus on the specification and development of pPGC, and therefore E14.soma.g1 and E14.soma.g2 are treated as one lineage (named as 'E14 Soma') in most of the analyses.

To identify the unique features of pPGC during pre-migratory stage and gonadal stage, DEG analyses were performed for E11 Epi vs. E14 PGC (Figure 2-13, 1151 up, 880 down in E14 PGC), E14 PGC vs. E14 Soma (Figure 2-14, 717 up, 1494 down in E14 PGC), and E14 PGC vs. E31 PGC (Figure 2-15, 136 down, 170 up in E31 PGC).

Enrichment analysis of GO biological process (BP) term and KEGG pathways were performed for DEG sets in each comparison. When comparing E14 PGC to E11 epiblast, the upregulated genes in E14 PGC are enriched for GO terms such as "positive regulation of differentiation" and "spermatogenesis, which is

consistent with E14 PGC being a differentiated cell type. Similarly, downregulated genes in E14 PGC show enrichment for "tissue morphogenesis", and "regulation of stem cell division", suggesting gradual loss of pluripotency in PGC (Figure 2-13).

A large number of genes show lower expression in E14 PGC compared to surrounding somatic cells. Those genes show enrichment for multiple embryonic development-related terms, including "multicellular organism development" and "anterior/posterior pattern specification". This is consistent with the repression of somatic lineage genes in PGC. On the other hand, genes with higher expression in E14 PGC show enrichment for germline-related term "germ cell development" (Figure 2-14).

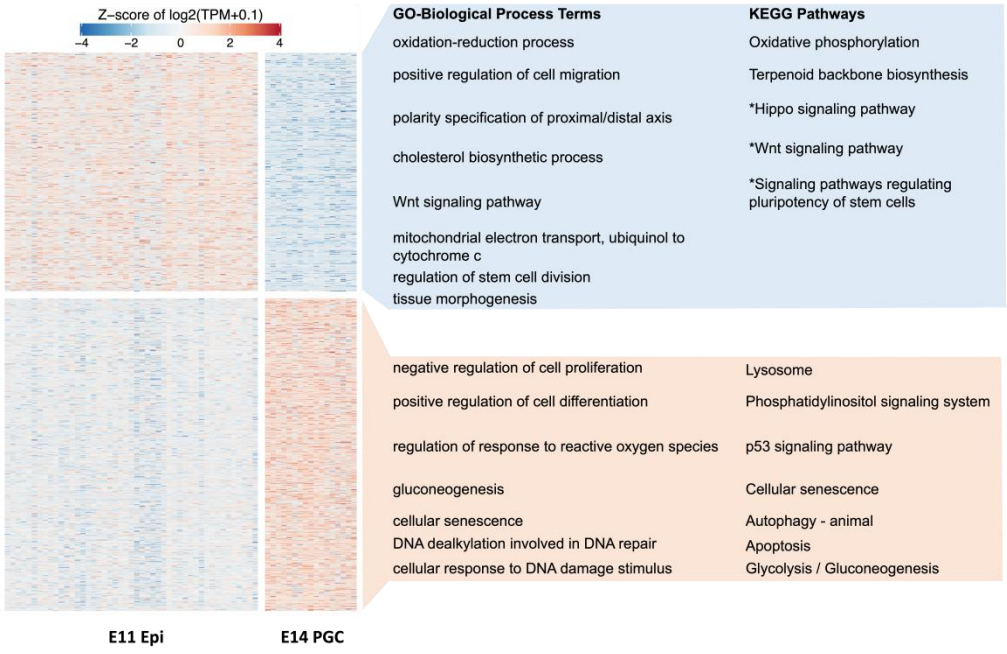


*Figure 2-13 Expression heatmap of DEGs between E14 PGC and E11 Epiblast with enriched GO terms for each cluster. * indicates the terms have significant p value ($p<0.05$) but not q value ($q>0.05$).*

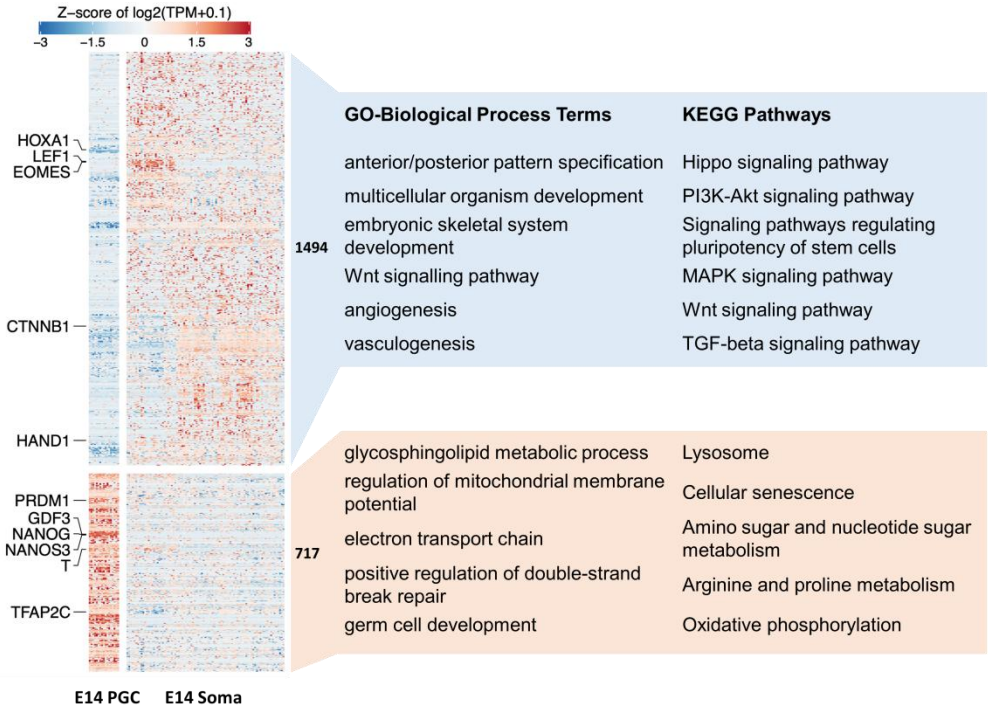


Figure 2-14 Expressional heatmap of DEGs between E14 PGC and E14 Soma with enriched GO terms and KEGG pathways for each cluster. A few lineage markers are indicated on the left side.

More germline-related terms, such as "piRNA metabolic process', are enriched with upregulated genes in E31 PGC, suggesting further development of the germline. Term "meiotic cell cycle" is also enriched for E31 PGC-upregulated genes (Figure 2-15). Although component of meiosis-specific cohesion complex, including *REC8, SMC1B* and *STAG3*, and synaptonemal complex proteins including *SYCP3* and *SYCE1*, are upregulated in some of the E31 pPGC, key meiotic recombination proteins such as the meiosis-specific nuclease *SPO11* are still absent, suggesting E31 female PGC has not entered meiosis prophase (Figure 2-16), which is consistent with previous observation that female pPGC begins to enter meiosis prophase around E47 (Bielanska-Osuchowska, 2006, Byskov et al., 1986). The upregulation of the abovementioned cohesion and synaptonemal components are observed in both male and female E31 PGCs, especially in a small group of male cells at the

right end of Figure 2-16. This suggests that preparation for meiosis may have already started in E31 PGC, regardless of the gender.

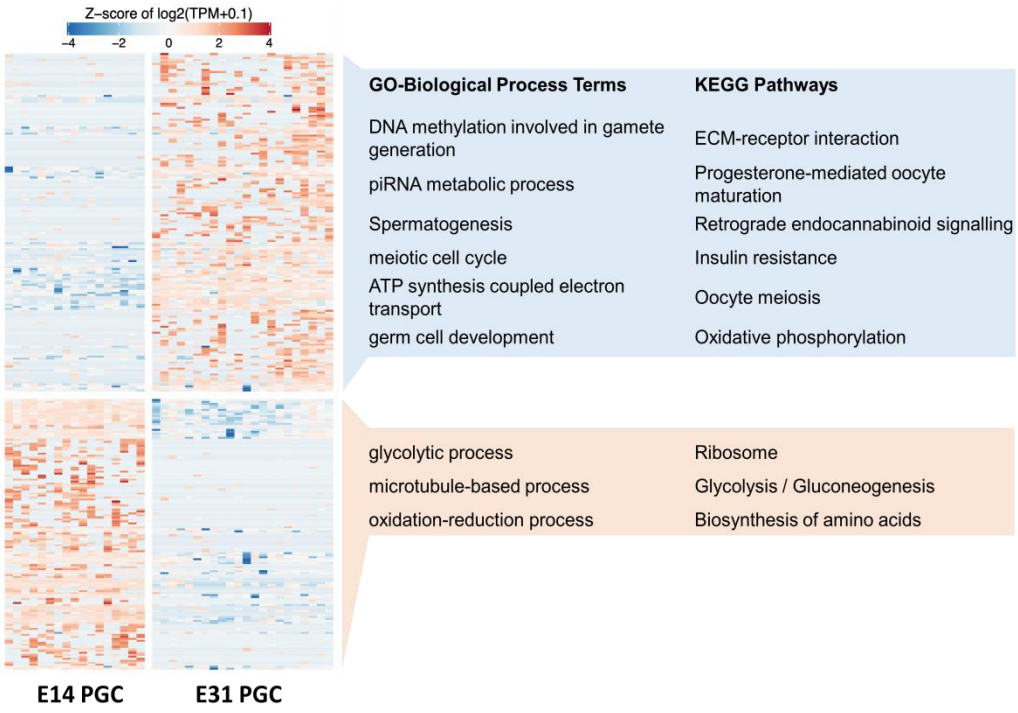


Figure 2-15 Expressional heatmap of DEG between E14 PGC and E31 PGC with enriched GO terms and KEGG pathways for each cluster.

Multiple terms related to DNA damage and repair, such as “positive regulation of double-strand break repair”, “cellular response to DNA damage stimulus”, “p53 signalling pathway” are enriched with E14 PGC-upregulated genes when compared to E11 Epiblast and E14 Soma (Figure 2-13, Figure 2-14). The enrichment of DNA repair terms in pre-migratory pPGC indicates that DNA repair mechanisms may be active immediately after PGC specification, which is consistent with a recent finding that DNA repair is critical for pre-meiotic PGC development (Hill and Crossan, 2019). Expression of DNA repair components is further discussed in section 2.3.2.1 . GO analysis of DEGs also reveal enrichment for multiple metabolism-related terms in E14 PGC when compared to E11 Epiblast and E14 Soma, which suggests metabolic changes such as reduced glycolysis and mitochondria (Figure 2-13, Figure 2-14). The

downregulation of glycolysis appears to further proceed in E31 PGC (Figure 2-15). These indicate an early and continuous metabolic shift in pPGC, which is further examined in section 2.3.1.7.

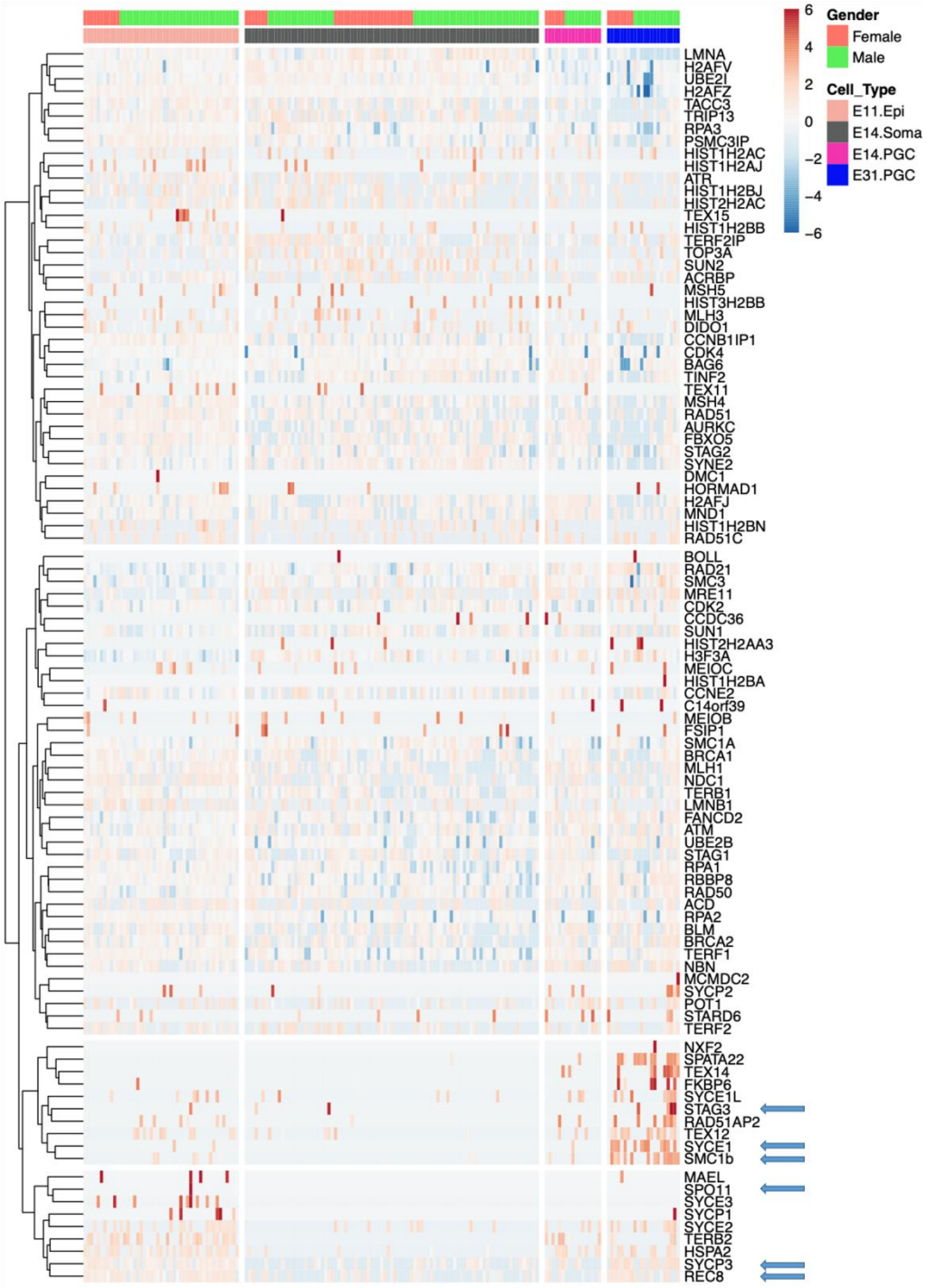


Figure 2-16 Expressional heatmap of meiosis-related genes selected based on GO term "Synapsis" (GO:0007129) and Reactome pathways "Meiotic synapsis" (R-HSA-1221632) and "Meiotic recombination" (R-HSA-912446).

The unit for colour scale is 'Z-score of log2 (TPM + 1)'. Blue arrow indicates genes mentioned in the text'

Moreover, the DEG results are compared with self-organizing maps (SOM) results (Figure 2-17) to identify cell type-specific genes. 1058, 442 and 145 genes that show differential expression pattern and reach highest average expression level (geometric mean) in E14 Soma, E14 pPGC and E31 pPGC, respectively, were identified as cell type-specific genes of the according stage. As expected, for E31 PGC-specific genes, GO terms analysis revealed enrichment for germline-specific terms such as “piRNA metabolic process”, “DNA methylation involved in gamete generation” and “spermatogenesis”. For E14 Soma-specific genes, general embryonic development terms such as “multicellular organism development” and “regulation of developmental process” were enriched (Figure 2-18). Surprisingly, enrichment analysis for E14 PGC-specific genes show only a few significantly enriched terms. This may be due to that most of germline-specific genes either do not show differential expression between E14 and E31 PGC, such as *BLIMP1* and *KIT*, or show higher expression in gonadal stage, such as *DAZL*. Similarly, features that shared between E14 PGC and E11 Epiblast are not included in this enrichment analysis.

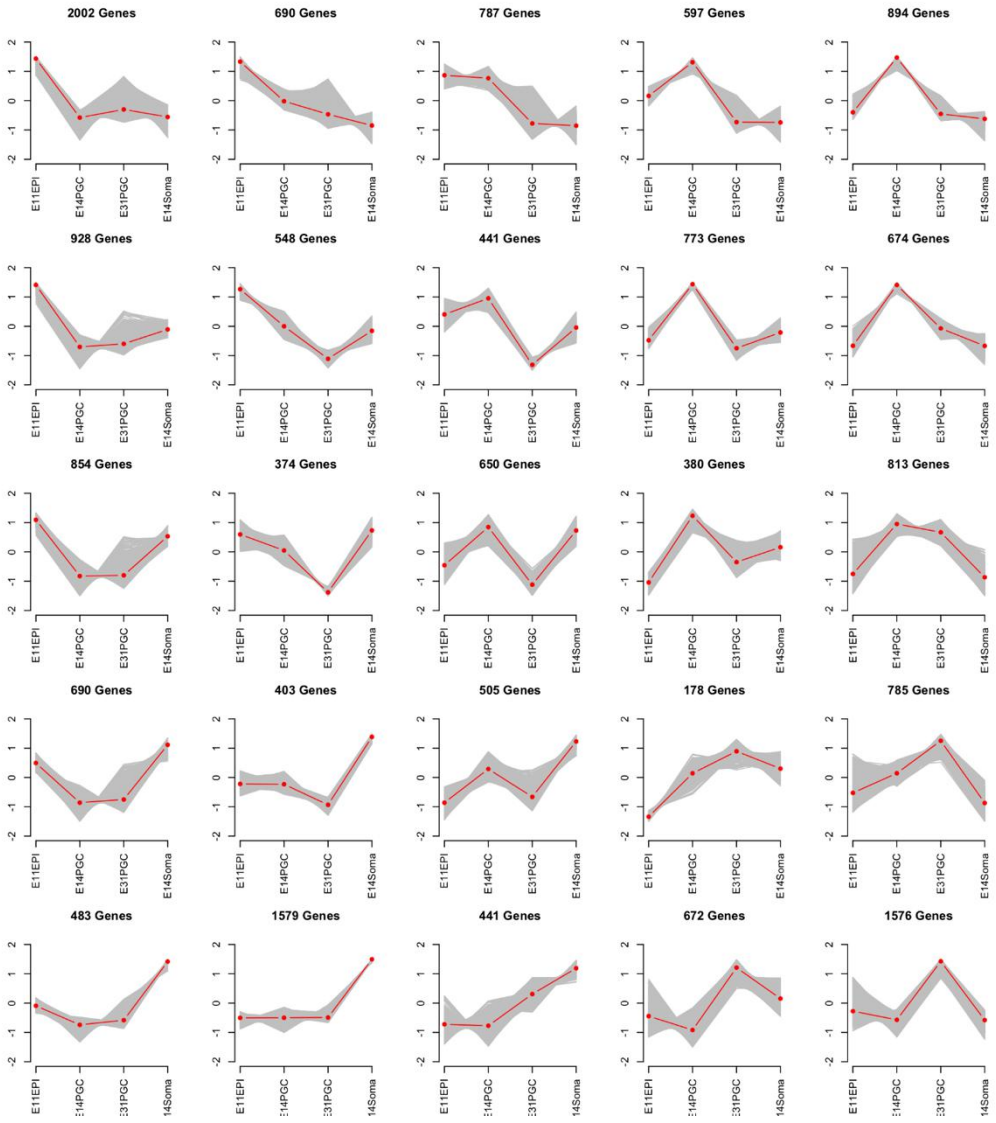


Figure 2-17 Line plot showing expression profiles of 25 clusters of genes classified by SOM analysis

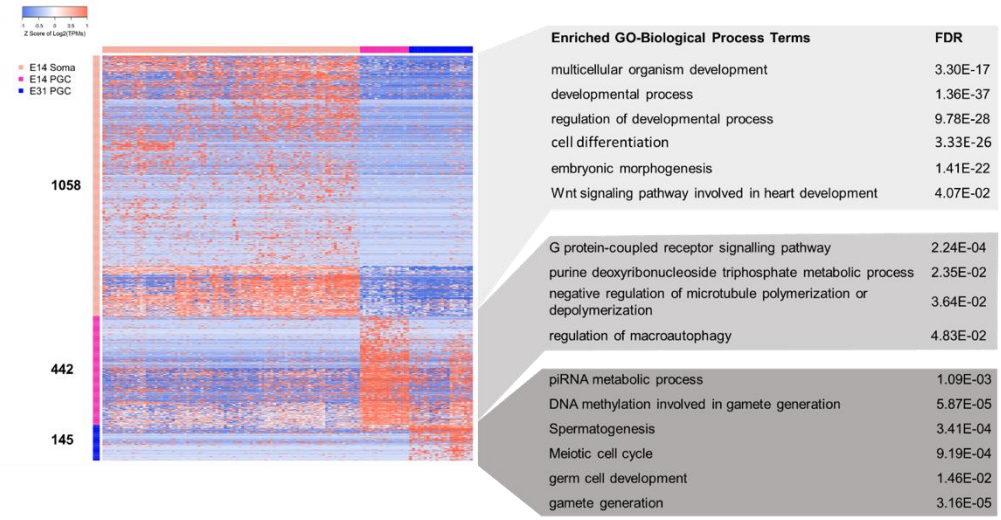


Figure 2-18 Dendrogram and expressional heatmap of cell type-specific genes with enriched GO terms.

In general, DEG identification and GO enrichment analysis reveal the distinction between pre-migratory pPGC and surrounding somatic cells, which suggests reprogramming events have begun immediately after the specification of PGC. Comparison between E14 and E31 PGC reveal the further development of germline, including preparation of meiosis and continuous changing at metabolic level.

2.3.1.4 Comparison to human and Cynomolgus monkey PGC reveal similar and unique features of mammalian PGC development

After confirming the identity of the PGC clusters, the expression profile of lineage markers was compared between hPGCs (Kobayashi et al., 2017), cyPGCs (Sasaki et al., 2016) and pPGCs. pPGC show similar profiles for key germline genes (e.g. *PRDM1, TFAP2C, NANOS3, DAZL*) and pluripotency genes (e.g. *NANOG, POU5F1, SOX2, ESRRB*) with the primate PGCs (Figure 2-19). Similar to hPGCs, *GATA4* is also widely expressed in pPGCs. *T (Brachyury)* is maintained in some gonadal pPGC, while *EOMES* is absent from pre-migratory stage. However, a few genes show different expression between two species. For example, naïve pluripotency gene *KLF4 and*

TFCP2L1 as well as trophectoderm marker *TEAD4,* which are expressed in hPGCs and hPGCLCs (Tang et al., 2015), are absent in most pPGCs. On the other hand, pluripotency genes *ZIC3* are expressed in early pPGCs, but not detected in hPGCs from migratory stage and downregulated in hPGCLCs (Tang et al., 2015, Kojima et al., 2017). *PRDM14*, a key germline gene in mPGC and hPGC development, is only detected in a few gonadal pPGCs and absent at pre-migratory stage. Hence, pPGCs and hPGCs share most of key genes for germline specification, including *SOX17*, BLIMP1 and TFAP2C, but show difference for *PRDM14* and some pluripotency genes.

As single-cell transcriptome data are available in hPGCs (Kobayashi et al., 2017) and cyPGCs (Sasaki et al., 2016), I compared PGCs and somatic lineages from these three species at transcriptomic level. In total, 14692 genes that are common among the datasets of three species (i.e. homologues genes with the same gene name in three species) were used for dimension reduction (UMAP) in *Seurat* after integration. The annotation of human and Cy monkey cell types is adopted from the original analysis of these datasets. Majority of pPGCs and cyPGCs cluster together with hPGCs, suggesting PGCs of three species show similar features and are distinct from somatic tissues. However, a large portion of E14 pig somatic cells (pSoma) and Cy gastrulating cells (cySoma) fail to cluster with human gonadal somatic cells, but instead cluster with PGCs (Figure 2-20). This might be due to the large difference between the less differentiated, multipotent cells in pig and Cy monkey, and the highly differentiated human gonadal somatic cells.

Indeed, when integrating only the pig and Cy monkey datasets, cells with similar identities in two species cluster nicely with each other. As early pluripotent lineages were available in both pig and Cy monkey dataset, these cell types were also included to support the functionality of the integration analysis. Pre-migratory pPGCs (E14) cluster closer with early (pre-gonadal) cyPGCs and gonadal pPGCs (E31) closer with late (i.e. gonadal) cyPGCs (Figure 2-21, left bottom). Majority of E14 pig somatic cells cluster with gastrulating cells (Cy Gast Soma). Note that one Cy monkey cell annotated as gastrulating cell clusters with PGCs, which is also observed in the original analysis (Sasaki et al., 2016). Like in other ungulates such as sheep and cattle

but unlike mouse and primates, pig has a longer period of pre-implantation stage. The implantation of pig embryo only begins around E12-13 and completes around E18 (Keys and King, 1989). Pig epiblast progresses through gastrulation before implantation of the conceptus (Talbot et al., 1993). Therefore, it is not surprising that the pre-implantation E11 pig epiblast cells, which is very close to gastrulating stage (E11.5-12) (Kobayashi et al., 2017), cluster with post-implantation Cy epiblast. Interestingly, pre-implantation Cy epiblast cells form a separate cluster that is close to morula and ICM but not overlapping with any pig cells, suggesting this specific cell type is not captured in the pig dataset, or not exist in pig.

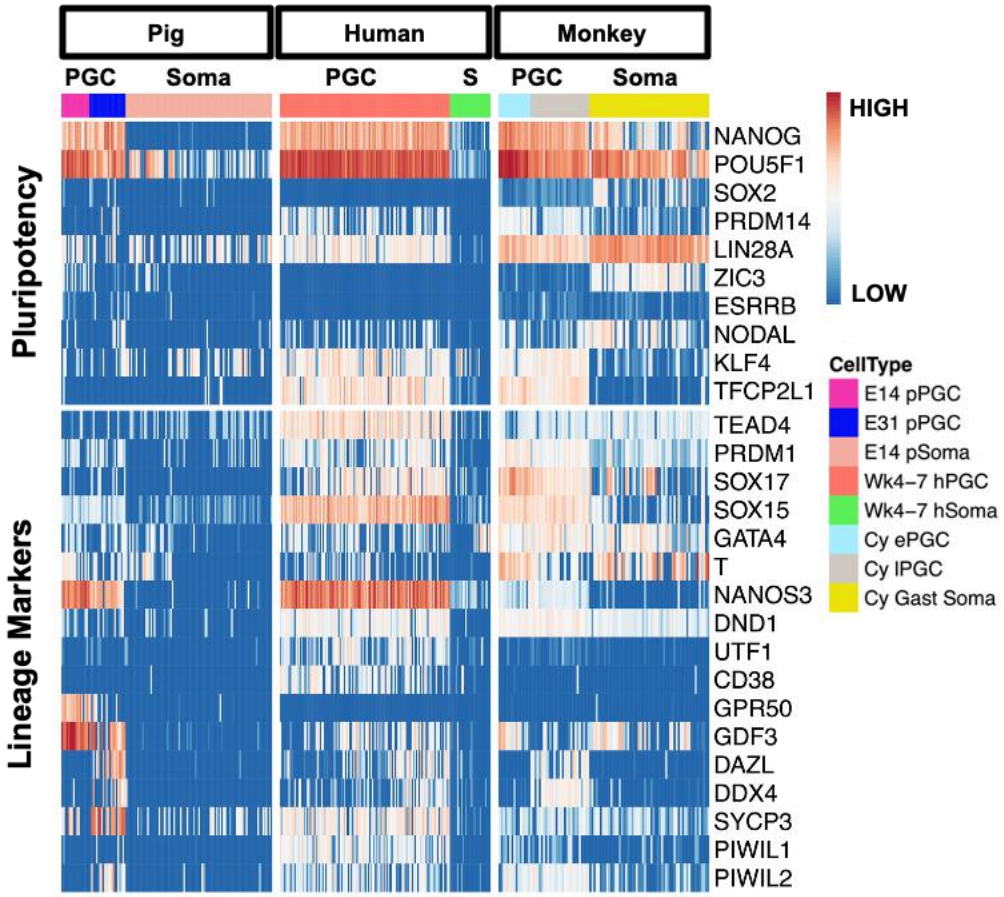


Figure 2-19 Expressional heatmap of PGC markers in listed samples. The human datasets from Lili et al. and Sasaki et al. are used in this plot. "S": Soma. "Cy ePGC": early cyPGC (E13-20); "Cy lPGC": late cyPGC (E36-55); "Cy Gast Soma": cy gastrulating cells (E13-20). As different expression units

are used for the three species, values in the colour scale are replaced by HIGH and LOW.

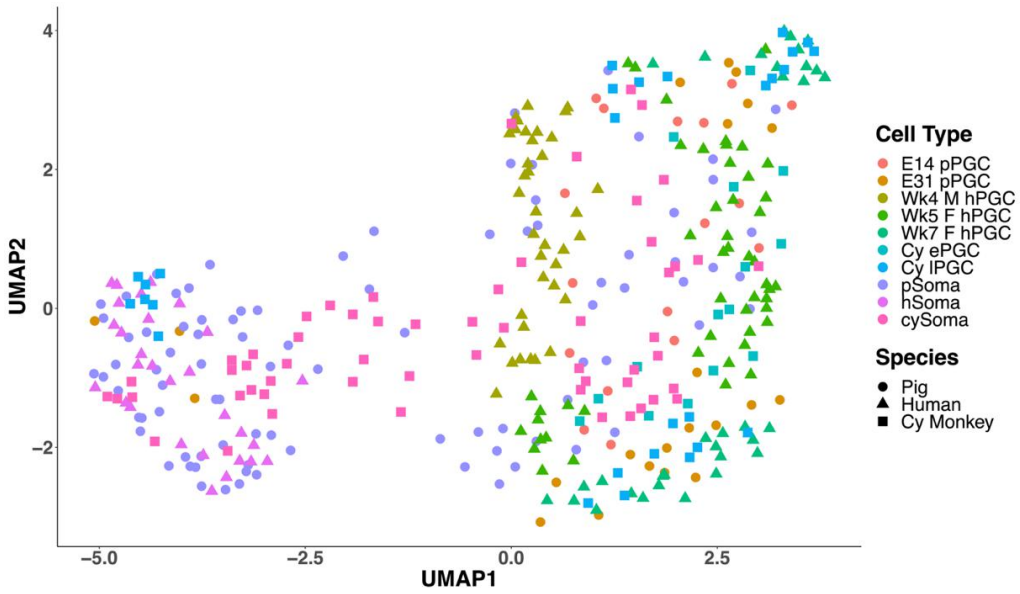


Figure 2-20 UMAP plot of integrated of human, Cy monkey and pig datasets. Majority of pPGCs and cyPGCs cluster with hPGCs, despite the difference in stages. "Cy ePGC": early cyPGCs (E13-20); "Cy lPGC": late cyPGCs (E36-55); "pSoma": E14 pig somatic cells; "hSoma": Wk4-7 human gonadal somatic cells; "cySoma": Cy monkey gastrulating cells (E13-20).

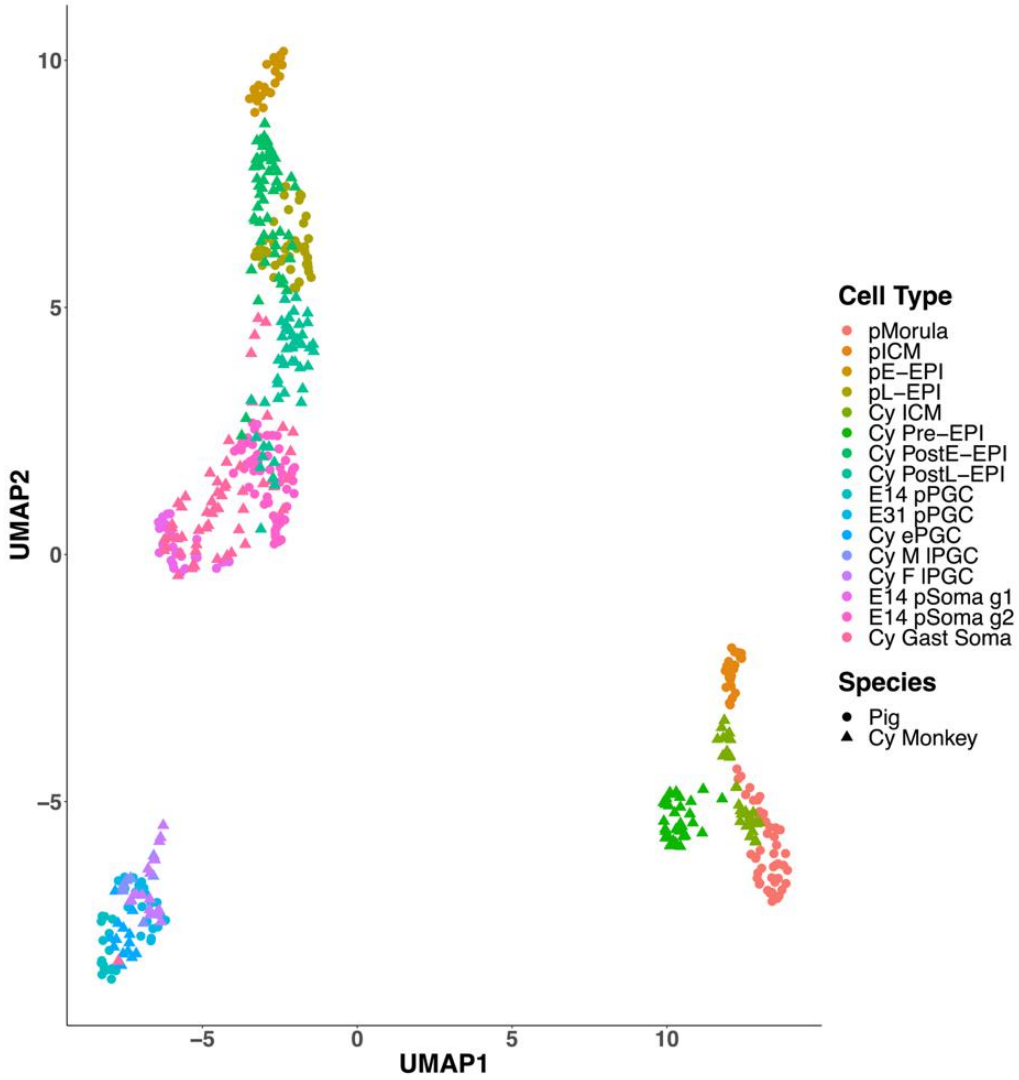


Figure 2-21 UMAP plot of integrated pig and Cy monkey datasets. Cells with similar identity in two species cluster nicely with each other. pMorula: pig morula (E5); pICM: pig ICM (E6); pE-EPI: pig early epiblast (E8); pL-EPI: pig late epiblast (E11); Cy ICM: Cy monkey ICM (E6); Cy Pre-EPI: pre-implantation epiblast (E7-9); Cy PostE-EPI, post-implantation early EPI (E13-14); Cy PostL-EPI, post-implantation late EPI (E16-17); Cy ePGC: early cyPGCs (E13-20); Cy M lPGC: late male cyPGCs (E36-55); Cy F lPGC: late female cyPGCs (E43-51); Cy Gast Soma: Cy monkey gastrulating cells (E13-20).

2.3.1.5 pPGC maintains partial features of pluripotency compared to somatic cells

Previous studies have detected features of both naïve and primed pluripotency in mPGCs and hPGCs, such as core pluripotency markers NANOG and OCT4 in both species, as well as naïve pluripotency markers *Esrrb* and *Nr5a2* in mPGC (Bell et al., 2020, Zhang et al., 2018b, Hackett et al., 2018) and *KLF4* and *TFCP2L1* in hPGC and cyPGC (Tang et al., 2015, Sasaki et al., 2016) (Figure 2-19). As shown in previous section (2.3.1.4), pPGCs express NANOG and OCT4, but do not show enrichment for *ESRRB*, *KLF4* or *TFCP2L1*. Therefore, I further investigate the pluripotency feature of pPGCs.

Similar to the investigation of amnion features in section 2.3.1.22.3.1.2.3, marker genes (i.e. genes with higher expression) in pig E6 ICM, E8 and E11 epiblast (Ramos-Ibeas et al., 2019), respectively, formed the signature sets for these pluripotent states. Expressions of the three signature sets were examined in pre-migratory (E14) and gonadal pPGCs (E31). Compared to E14 somatic cells, a strong pig E8 Epi signature score was determined for both pPGC stages, with gonadal pPGCs also showing strong ICM signature score (Figure 2-22).

The signature genes contributing to these scores include elevated expression of well-known transcription factors (*POU5F1, NR5A2, SOX15* and *OTX2*), and also of chromatin related genes (*HELLS, BRDT, ZAR1*) and repressors of TE activity (*MOV10, ASZ1, PLD6, HENMT, TDRKH* and *SAMHD1*), indicating that restoration of a gene signature common with ICM/E8 Epi in the early PGCs is linked to epigenetic resetting in the germline (Figure 2-23).

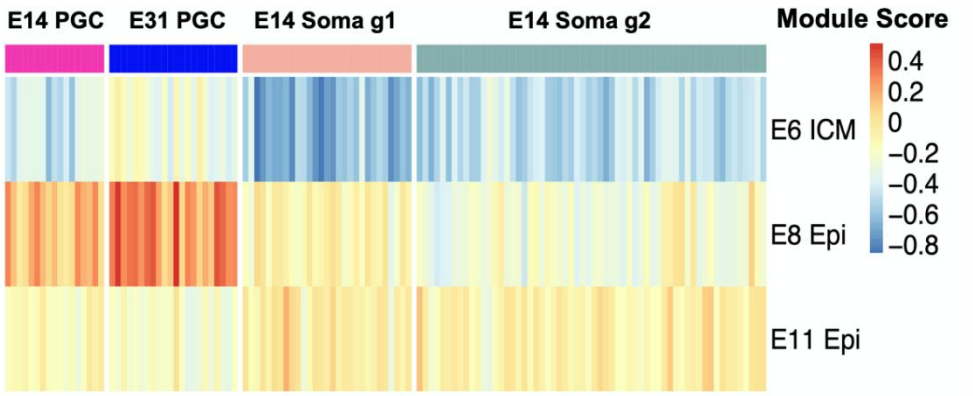


Figure 2-22 Relative average expression (module score) of signature sets of E6, E8 and E11 cells in E14 and E31 cells. E31 PGCs show enriched expression of E6 ICM and E8 Epi signature sets.

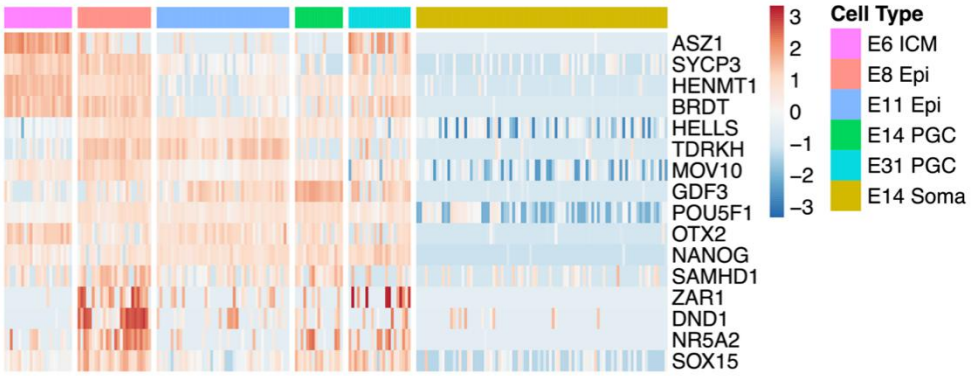


Figure 2-23 Scaled expression (Z-score of log2(TPM+1)) of selected genes that are highly expressed in both pig pluripotent cells (E6-11) and PGCs.

2.3.1.6 Novel membrane proteins identified in pPGC

Plasma membrane proteins participate in numerous cellular processes, such as signalling, transport and cell migration (Tan et al., 2008, Calabrese and Radford, 2018). Therefore, I sought to identify potential pPGC-specific plasma membrane proteins, by selecting pPGC-specific genes with relevant GO terms ("Cell surface" GO:0009986 and "Plasma membrane" GO:0005886) and/or curated in the Cell Surface Protein Atlas (Bausch-Fluck et al., 2015). Similar to previous reports in hPGCs and cyPGC (Tang et al., 2015, Gomes Fernandes

et al., 2018, Sasaki et al., 2016), surface markers *KIT* and *PDPN* are upregulated in pre-migratory pPGCs (Figure 2-24 and Figure 2-26). An orphan receptor, GPR50, is specific for early but not gonadal pPGCs. Although by IF GPR50 is detected in presumably the nucleoli in all cells and stages, only early migratory pPGC shows GPR50 immunoreactivity at cell membrane, which is also stronger than the nucleoli staining (Figure 2-25). E14 pPGC also show high level of CXCR4 (Figure 2-2), which suggest the imminent onset of migration (Molyneaux et al., 2003, Mitsunaga et al., 2017). GDF3, a mammalian-specific TGFβ ligand, which inhibits BMP/TGF-beta signalling (Levine and Brivanlou, 2006), also show enrichment in early pPGC, cyPGC (Sasaki et al., 2016) and gonadal hPGC. Upregulation of vitamin C transporter *SLC23A2* is also observed from pre-migratory stage in pPGC and cyPGC and in gonadal hPGC, which may increase the cellular uptake of vitamin C and promote TET1 activity in PGC (DiTroia et al., 2019) (Figure 2-24 and Figure 2-26). A few CD markers, such as CD126 (IL-6R), CD157 (BST1) which is a paralogue of hPGC marker CD38, and another orphan receptor GPR133 (*ADGRD1*), are also upregulated in pPGC (Figure 2-24), suggesting a unique membrane protein signature potentially critical for PGC development and function.

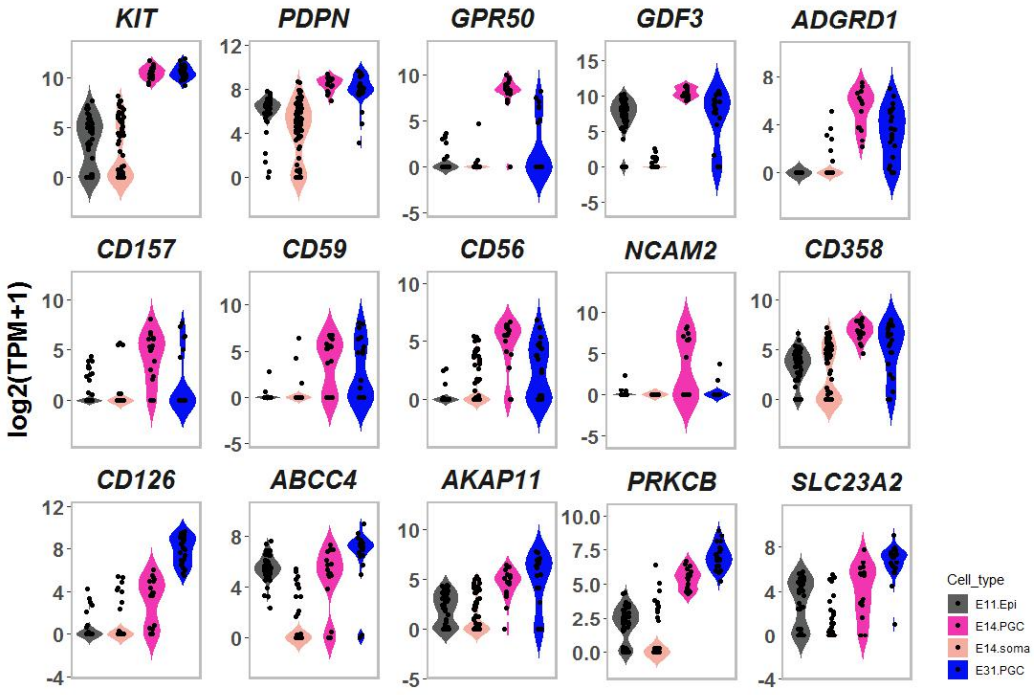


Figure 2-24 Violin plot of the expression of selected plasma membrane protein in listed samples.

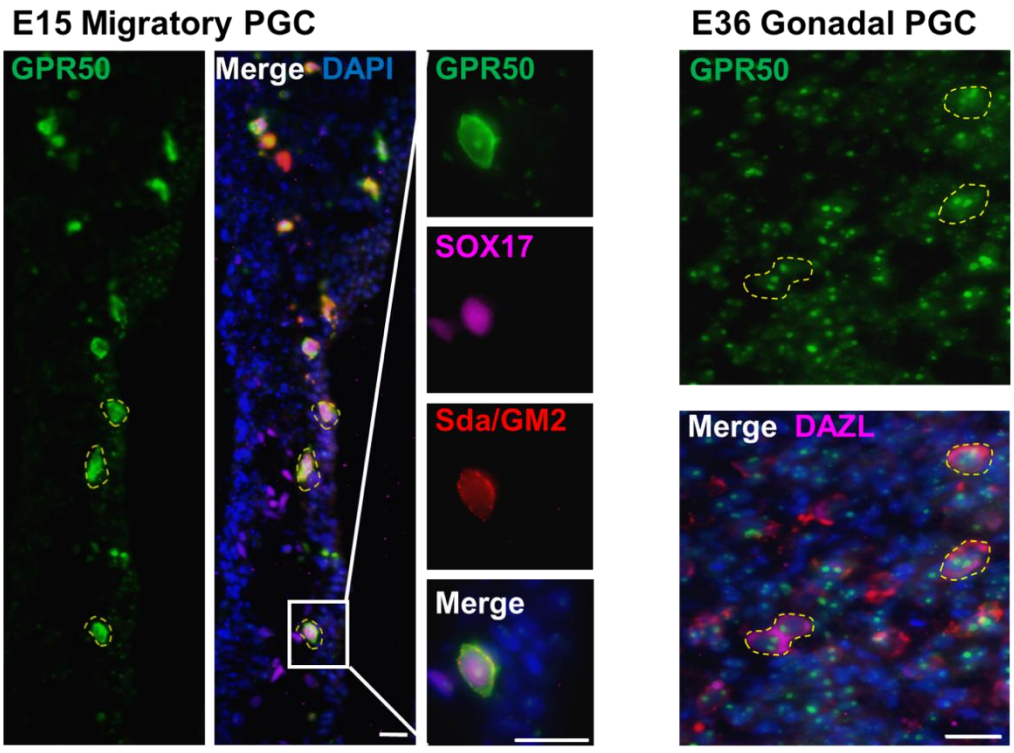


Figure 2-25 Immunofluorescence of GPR50 in E15 early migratory pPGC (left) and E36 gonadal pPGC (right). GPR50 shows membrane immunoreactivity and co-stains with PGC markers SOX17 and Sda/GM2 in E15 pPGCs, but does not stain cell membrane or show specificity towards pPGCs in E36 gonad.

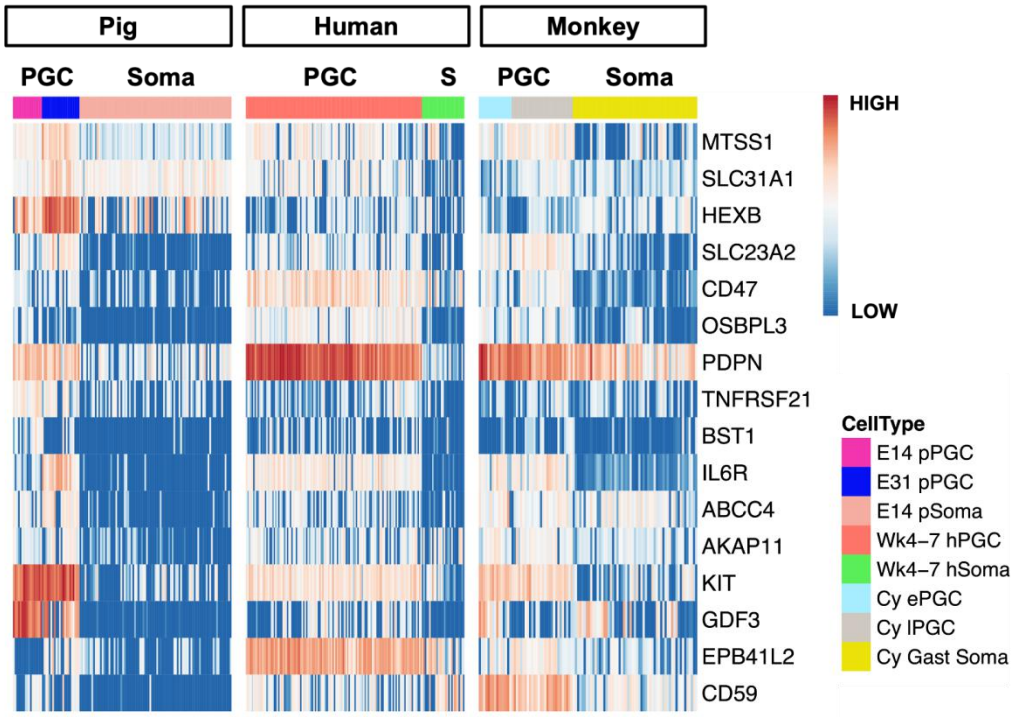


Figure 2-26 Expressional heatmap of selected plasma membrane protein in human, Cy monkey and pig PGC and somatic cells. Known PGC surface markers PDPN and KIT, as well as other membrane proteins such as Vitamin C transporter SLC23A2, show enriched expression in PGCs of three species.

2.3.1.7 Expressional changes of metabolic-related genes in pPGC

Metabolic changes have been observed in various processes during cell fate change, including exit from and reprogramming to pluripotency, suggesting a tight link between metabolism and cellular identity. Recent studies have shown that metabolic shift occurs during hPGC (Floros et al., 2018) and mPGC development (Hayashi et al., 2017) and supports hPGCLC induction (Tischler et al., 2019, Chen and Clark, 2018). The GO enrichment analysis for DEGs

between different cell types also revealed multiple metabolism related terms, including downregulation of “glycolytic process” and upregulation of “ATP synthesis coupled electron transport” in E31 PGC (Figure 2-15). Based on GO enrichment results, expression of components of glycolysis, tricarboxylic acid (TCA) cycle, oxidative phosphorylation and cholesterol metabolism are further investigated here.

Consistent with the findings in E13.5 mPGC, gonadal hPGC and hPGCLC, glycolysis is largely downregulated in E31 gonadal pPGC, compared to all other samples. Most enzymes that contribute to glycolysis, including the rate-limiting phosphofructokinase (e.g. *PFKL*), are downregulated in E31 pPGC compared to E14 pPGC (Figure 2-27). Surprisingly, many of these glycolytic-related genes (e.g. *GPI*, *PFKL*, *ALDOA* and *GAPDH*) are already downregulated in E14 pPGC compared to their surrounding somatic cells, suggesting that downregulation of glycolysis begins in pre-migratory pPGC. On the other hand, fructose 1,6-bisphosphatase (*FBP1*) which mediates the rate-limiting step of gluconeogenesis, is largely upregulated in pPGC from pre-migratory stage.

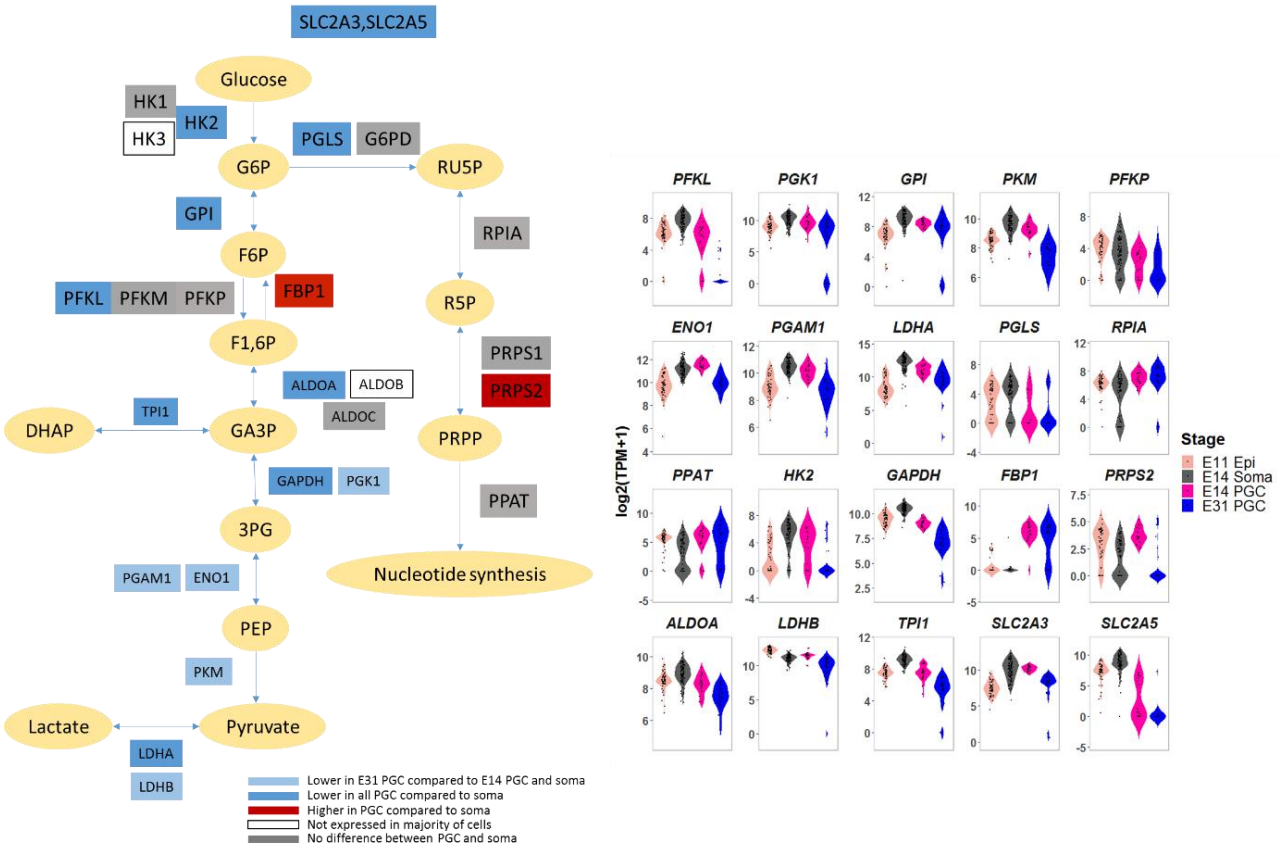


Figure 2-27 (Left) Illustration comparing glycolysis-related pathways between pPGC and somatic cells. (Right) Violin plot showing expression profile of selected glycolysis-related genes.

Concurrent with the downregulation of glycolysis, some components of the TCA cycle are upregulated in pPGC from pre-migratory stage, such as mitochondrial pyruvate carriers 2 (*MCP2*) which is responsible for transport of pyruvate into mitochondria matrix (Bricker et al., 2012), and dihydrolipoamide dehydrogenase (*DLD*) which functions as the E3 component of pyruvate and alpha-ketoglutarate (αKG) dehydrogenase complexes (Patel and Roche, 1990). The downregulation of pyruvate-dependent kinases 1 (*PDK1*), which phosphorylate the pyruvate dehydrogenase complex and reduce its activity (Patel and Korotchkina, 2001), may also contribute to the enhanced TCA cycle activity in pPGC (Figure 2-28). Although the reduced glycolytic activity and enhanced TCA cycle activity observed in pPGC is consistent with the metabolic shift from glycolysis to oxidative phosphorylation previously reported in mPGC and hPGCLC, the detailed mechanism may vary between species. For example, the upregulation of *IDH2*, which is reported in E13.5 mPGC and hPGCLC, is not detected in pPGC.

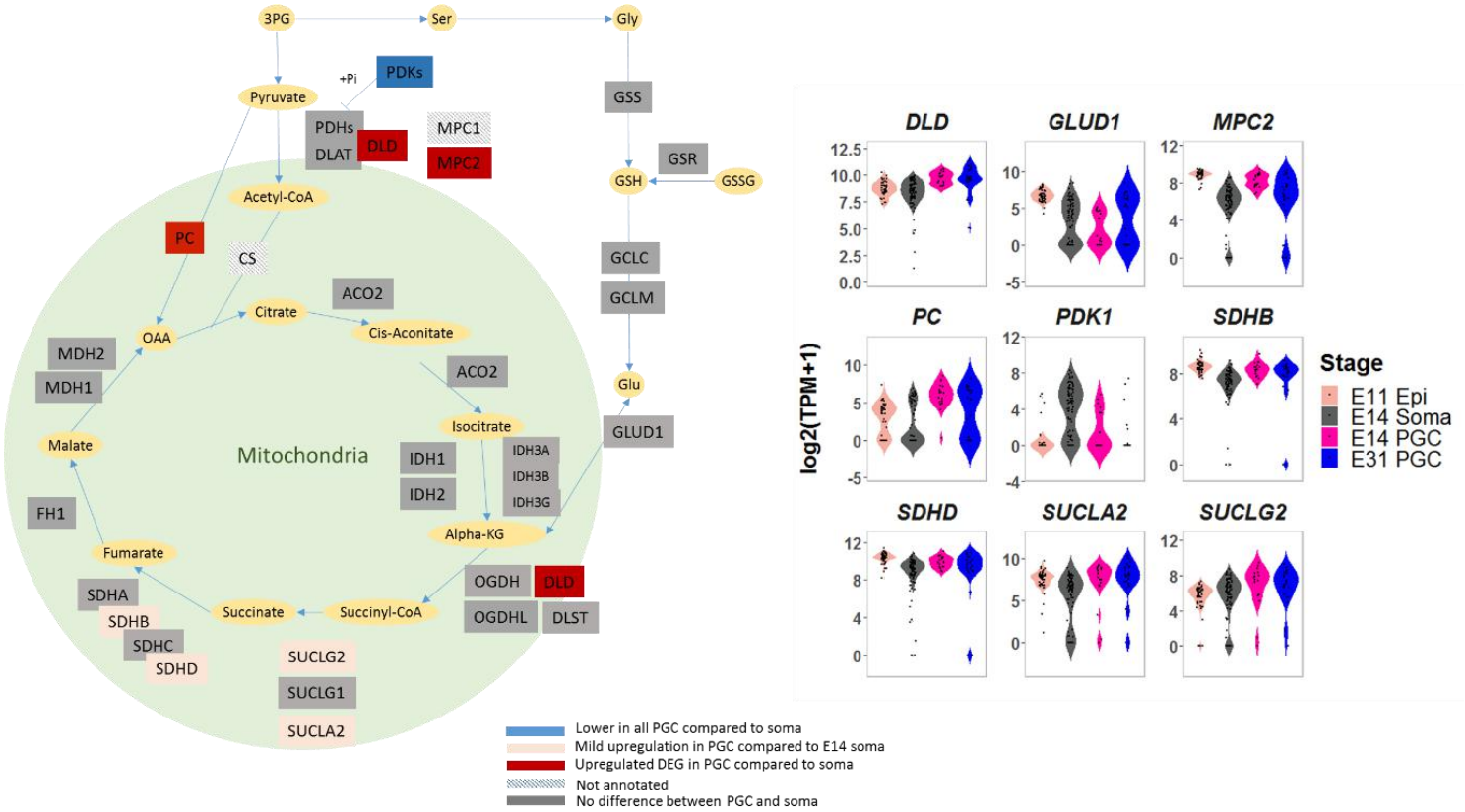


Figure 2-28 (Left) Illustration comparing TCA cycle-related pathways between pPGC and somatic cells. (Right) Violin plot showing expression profile of selected TCA cycle-related genes.

Surprisingly, the expression profile of mitochondrial electron transport chain (ETC), however, show downregulated expression of multiple components in E14 pPGC. Ten out of 13 mtDNA-encoded protein-coding genes, which are important components of ETC, are downregulated in E14 pPGC compared to E14 Soma (Figure 2-29). Furthermore, all 7 mtDNA-encoded protein-coding genes (*MT-ND1* to *MT-ND6*) in NADH dehydrogenase complex (i.e. Complex I of the ETC) show reduced expression in E14 pPGC. This suggests potentially supressed activity of ETC. However, whether the transcription from mtDNA is downregulated in E14 pPGC is unclear. Key components of mtDNA transcription machinery and potential nuclear regulation mechanisms of mtDNA transcription, such as the mtDNA-directed RNA polymerase (*POLRMT*), do not show consistent downregulation in E14 pPGC (Figure 2-30) (D'Souza and Minczuk, 2018, Kumar et al., 2018, Ali et al., 2019). It is possible that the reduced expression of mtDNA-encoded genes is due to the lower mtDNA copy number in early pPGC, as a previous study suggests mtDNA copy number was reduced largely in both female and male hPGCLCs (Floros et al., 2018). Interestingly, mtDNA polymerase POLG, which is responsible for mtDNA replication and repair (Krasich and Copeland, 2017), shows increased expression in pPGC from pre-migratory stage (Figure 2-29). Consistent with the finding in hPGC, this suggest that mtDNA copy number may reach the lowest level in E14 pPGC and starts increasing afterwards (Chen and Clark, 2018).

Apart from the shift in energy metabolism, changes are also observed in cholesterol and fatty acid metabolism of pPGC. Multiple components of cholesterol biosynthesis, including the master transcriptional activator *SREBP2*, the rate-limiting enzyme HMG-CoA reductase (*HMGCR*) and cholesterol uptaker *LDLR* (Madison, 2016), are largely downregulated in pPGC from pre-migratory stage, while components of cholesteryl esterification including *LIPA* and *SOAT1* are upregulated (Figure 2-31). Some cholesterol biosynthesis enzymes, such as *DHCR7* which mediates the last step of cholesterol synthesis, show downregulation in E31 PGC compared to E14 PGC. Potentially affected by the repression of SREBP2, fatty acid synthase (*FASN*) is also largely absent

in pPGC. Interestingly, the repression of SREBP2 may also contribute to the downregulation of glycolysis, as a recent study reported that knockdown of SREBP2 reduced glycolytic potential in cancer cells (Wen et al., 2018). Overall, these observations suggest major downregulation of cholesterol and fatty acid synthesis in pPGC, potentially regulated via repression of SREBP2.

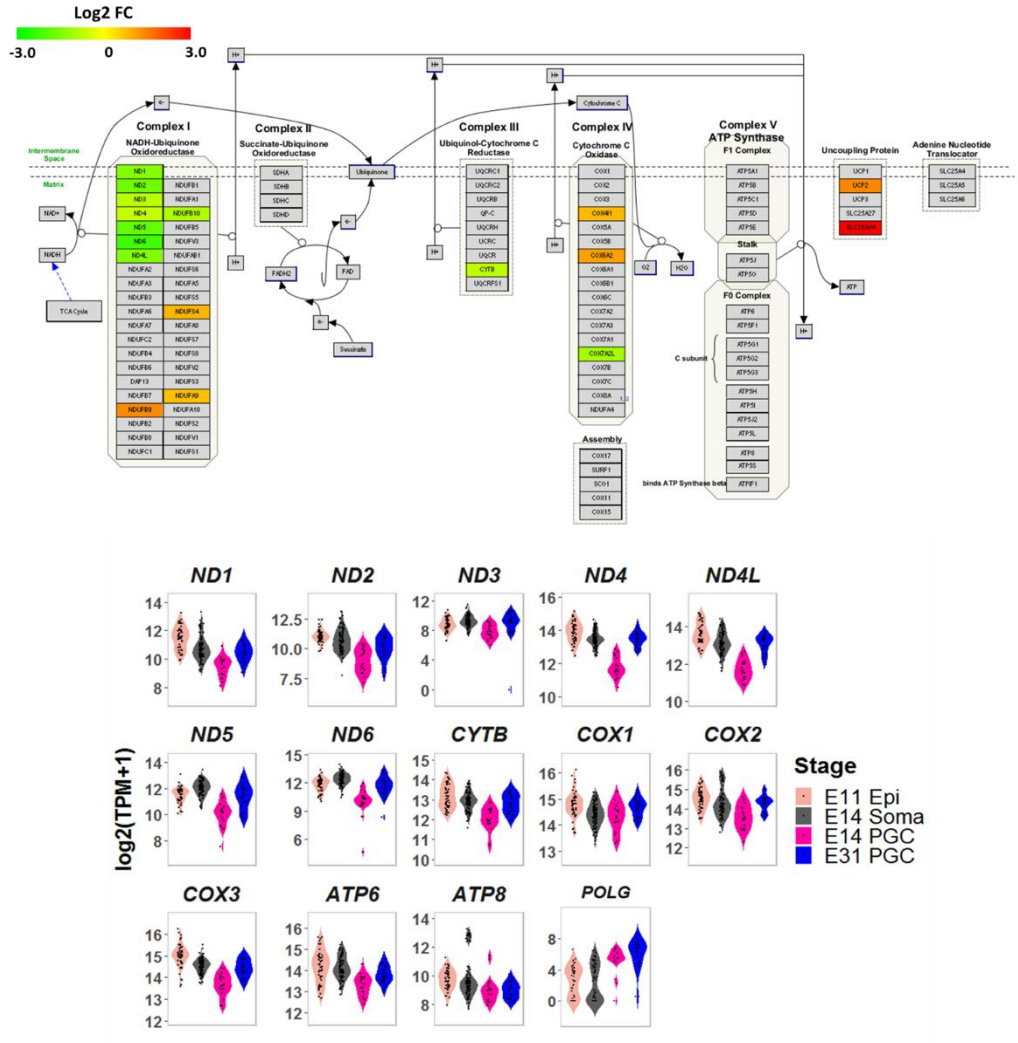


Figure 2-29 (Top) Pathway view comparing ETC-related genes between E14 pPGC and E14 somatic cells. Wikipathway "Electron Transport Chain (OxPhos system in mitochondria)" (WP111) is used as the template (Slenter et al., 2018). (Bottom) Violin plot showing expression profile of mtDNA-encoded protein-coding genes and mtDNA polymerase in listed samples.

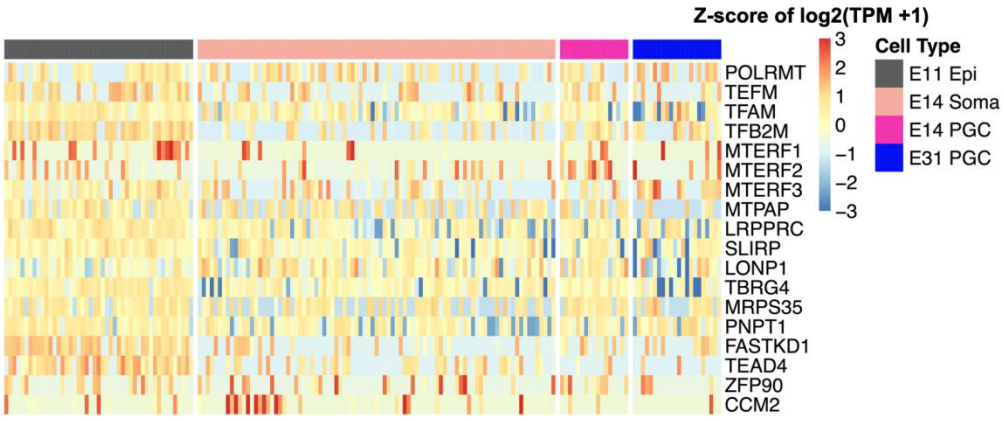


Figure 2-30 Expression of components and regulatory factors of mtDNA transcription. No consistent downregulation is observed in E14 PGC.

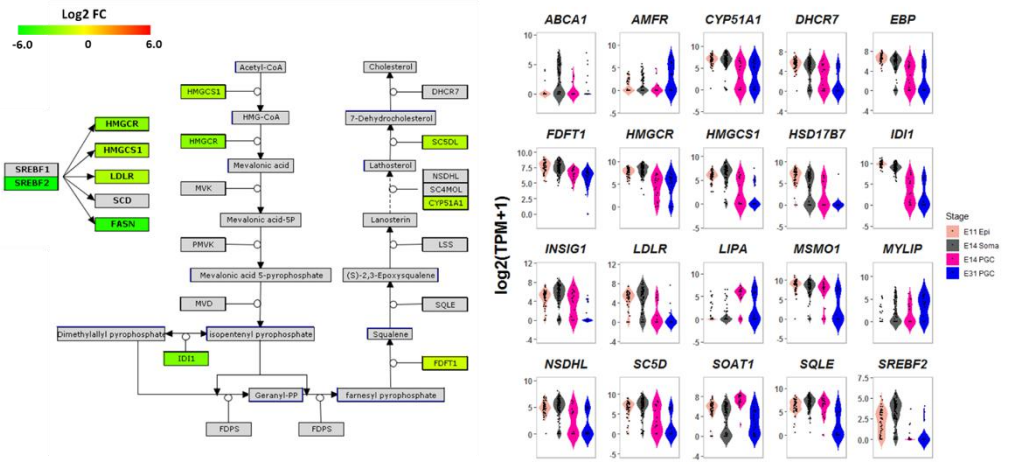


Figure 2-31 (Left) Pathway view comparing cholesterol synthesis-related genes between E14 pPGC and E14 somatic cells. The pathway template is modified from Wikipathway "Cholesterol Biosynthesis Pathway" (WP197) and "SREBF and miR33 in cholesterol and lipid homeostasis" (WP2011) (Slenter et al., 2018). (Right) Violin plot showing expression profile of selected genes related to cholesterol synthesis.

As mentioned above, metabolism related DEGs identified between pPGC and pig Soma show very few overlaps with metabolic components that are differentially expressed in E13.5 mPGC (Hayashi et al., 2017) or gonadal hPGC (Floros et al., 2018). This might because that the PGCs in three species were compared to different somatic lineages (i.e. early, gastrulating somatic

cells in pig but gonadal soma in human and mouse). Therefore, single cell dataset of Cy monkey was also included. Indeed, more metabolic components are found with similar expression profiles in pig and Cy monkey PGCs and early somatic lineages (Figure 2-32), further supporting the downregulation of glycolysis and cholesterol synthesis occur in early, pre-gonadal PGCs.

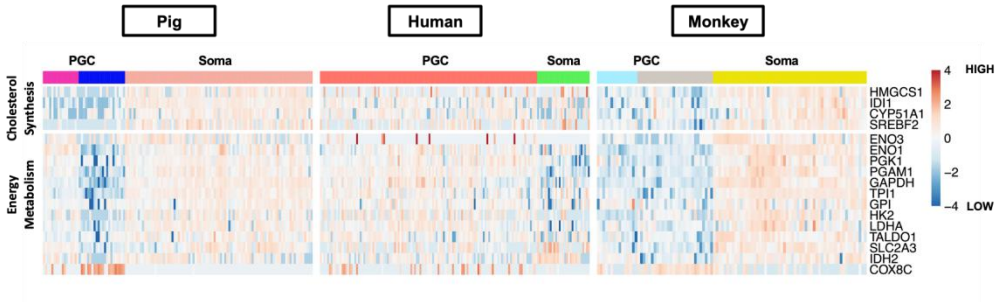


Figure 2-32 Expression profiles of selected metabolism genes in pig, human and Cy monkey. The profiles in pPGCs show large similarity with the one in cyPGCs, suggesting the dynamic changes, such as downregulation of glycolysis and cholesterol synthesis, may also occur in early and gonadal cyPGCs. Noting that as mtDNA-encoded genes were excluded in human and cy monkey datasets, those genes are not included in the comparison.

2.3.2 Epigenetic reprogramming events occur early in pre-migratory pPGCs

2.3.2.1 Active DNA demethylation is evident in pre-migratory pPGCs

DNA demethylation is a featured event during PGC development, which occurs in both passive (i.e. through dilution during DNA replication) and active (i.e. TET-mediated) way. In contrast to the active DNA demethylation only being detected in gonadal stage of mPGCs, it has been previously showed that the upregulation of *TET1* and its oxidized product 5hmC are detected in migratory hPGCs (Wk4). Partial erasure of 5mC coupled with increased level of 5hmC at H19 and GNAS ICRs is also reported in D4 hPGCLCs (Tang et al., 2015). Similarly, enrichment of 5hmC is detected in pre-migratory pPGC cluster (Figure 2-33 (A)) (Kobayashi et al., 2017). These suggest that active

DNA demethylation may occur at pre-gonadal stage in human and non-rodent PGC.

As IF results are only semi-quantifiable, liquid chromatography–tandem mass spectrometry (LC-MS/MS) was used for more accurate quantification of 5mC and 5hmC level (Hill et al., 2018). 5hmC level in FACS-sorted pre-migratory (E14) pPGC is indeed higher than the surrounding somatic cells and also the pig epiblast, suggesting active DNA demethylation is highly effective and the enrichment of 5hmC is unique in newly specified pPGCs. Concurrently, 5mC level drops in pre-migratory pPGCs (Figure 2-33 (C)). Consistent with IF results (Figure 2-33 (B)), 5hmC and 5mC are almost absent in FACS-sorted Wk5 gonadal pPGC, as previously suggested in gonadal hPGC (Guo et al., 2015, Tang et al., 2015). Similar trends of 5mC and 5hmC are observed in equivalent human samples, including the slight upregulation of 5hmC in D4 hPGCLC (Figure 2-33 (C)). Although the dynamics of 5hmC is unknown in cyPGCs, 5mC level is shown to reduce in early cyPGCs (E17) compared to surrounding somatic cells (Sasaki et al., 2016).

I further looked into the mechanism of DNA demethylation in pPGC. Although *DNMT1* is still present, which was also observed in hPGC (Guo et al., 2015), DNA methylation machinery is likely to be supressed as *UHRF1* and *DNMT3B* are absent in pPGC from pre-migratory stage (Figure 2-34 (A)). Surprisingly, despite the evident enrichment of 5hmC, no *TET* gene is upregulated transcriptionally in pPGC (Figure 2-34 (A)), suggesting the enrichment of 5hmC may be due to post-transcriptional regulation of TET genes and absence of DNA replication in pre-migratory pPGC (Kobayashi et al., 2017). Also, *DNMT3A* does not show transcriptional reduction in pPGC until gonadal stage (Figure 2-34 (A)). A previous study indicates that immunofluorescence of DNMT3A show reduction from migratory stage (E16) in pPGC (Kobayashi et al., 2017). Therefore, it is possible that DNMT3A still mediate de novo methylation in pre-migratory pPGC, while TET1 oxidizes the newly formed 5mC into 5hmC. DNA methylation machinery (i.e. *DNMT3A, DNMT3B, DNMT1* and *UHRF1*) is likely to be also repressed in early cyPGC.

Surprisingly, the catalytically inactive *DNMT3L* is enriched in early and late cyPGCs (Figure 2-35). DNMT3L has been shown to maintain DNMT3A stability (Veland et al., 2019) and stimulate its catalytic activity (Chédin et al., 2002), however IF in cyPGC suggests DNMT3A is almost absent from E17 (Sasaki et al., 2016). Therefore, the function of DNMT3L in cyPGCs remain unclear. *TET* genes appear to be upregulated in late cyPGCs compared to the early ones, consistent with upregulation of TET1 in gonadal hPGC. It remains to be investigated whether this early active DNA demethylation observed in pre-migratory pPGC also occurs in primate PGCs.

The removal of TET-oxidized products in PGC is suggested to be mediated by passive demethylation, i.e. dilution through DNA replication, as well as more dynamic mechanisms such as base excision repair (BER) pathway (Hill et al., 2018, Hackett et al., 2013, Hajkova et al., 2010). As DNA replication, and therefore passive demethylation, is largely absent in pre-migratory pPGC based on Edu staining (Kobayashi et al., 2017), BER pathway might be activated and contribute to DNA demethylation in pPGC from pre-migratory stage. Components of BER pathway, including *LIG1, POLD2, PARP1 and UNG*, are upregulated in E14 and E31 pPGC (Figure 2-34 (A)). The enrichment of PARP1 and UNG in early migratory and gonadal pPGC is also shown by IF (Figure 2-34 (B)). *TDG*, which was suggested to recognize TET-oxidative products and initiate BER pathway, is not differentially expressed in pPGC (He et al., 2011, Shen et al., 2013, Tang et al., 2015, Hajkova et al., 2010, Maiti and Drohat, 2011). Upregulation of some 'readers' for TET-oxidized products (*HELLS, HMCES, NUP133 and URB2*) is also observed in pPGC, suggesting 5hmC may be a dynamic and functional marker in PGC. A recent study indicates important role Fanconi Anaemia (FA) pathway in pre-meiotic mPGCs (Hill and Crossan, 2019). Although most of the DNA repair genes reported by Hill and Crossan are not differentially expressed in pPGC, upregulation of some FA components (*FANCI and FANCD2),* and a few other DNA-repair genes such as *PMS2 and NBN*, in E14 pPGC, suggesting multiple DNA repair mechanisms may be activated from pre-migratory stage in PGC. Overall, these results suggest extensive active DNA demethylation and DNA

repair occur in non-replicative pre-migratory stage of pPGC. Many of those DNA 'readers' and repair genes, such as *PARP1, UNG, HMCES, HELLS* and *FANCD2*, also show upregulation in gonadal hPGCs. However, most of them, apart from *UNG*, do not show clear upregulation in early (E13-20) cyPGCs (Figure 2-35), thus the functions of BER and other DNA repair pathways in early PGCs remain to be determined.

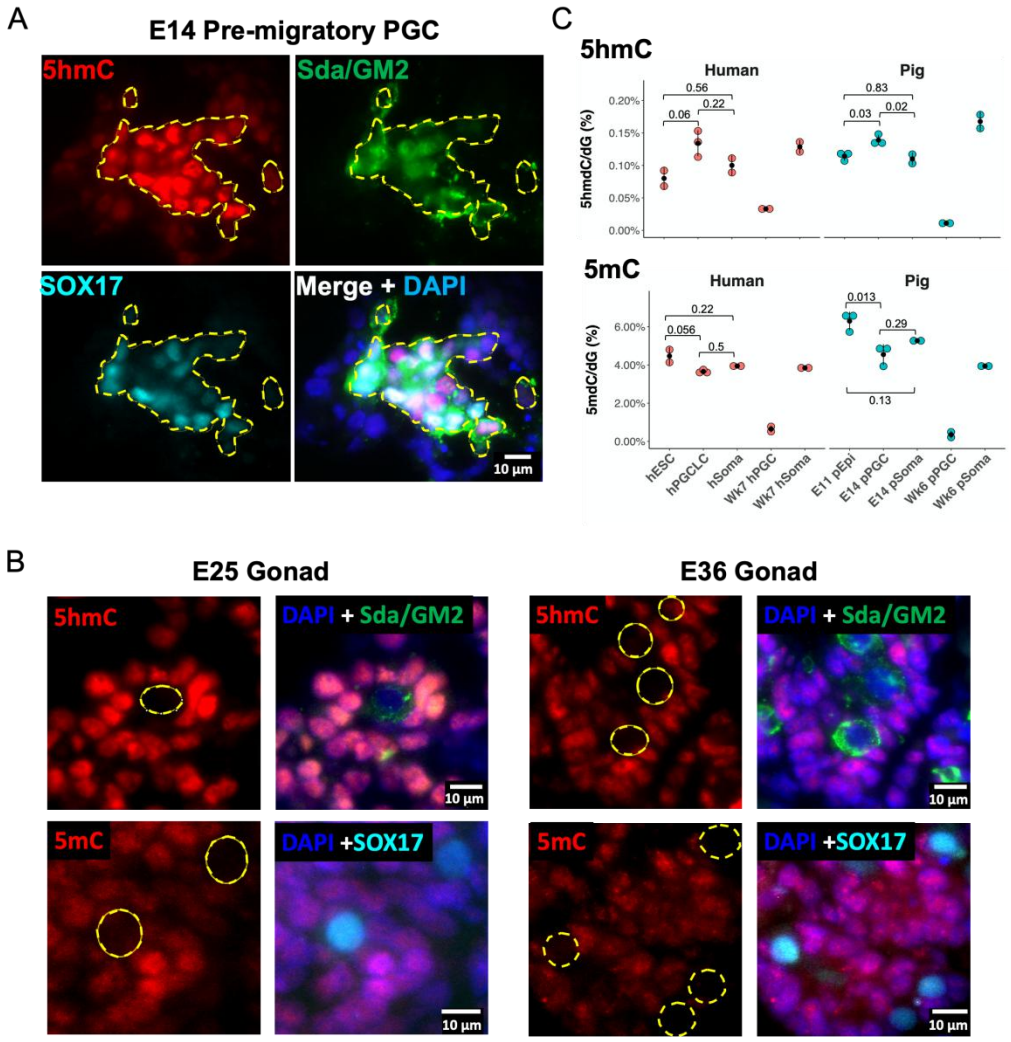


Figure 2-33 Immunofluorescence of 5hmC in (A) E14 PGC cluster and (B) 5mC and 5hmC in gonad sections of E25 and E36 embryos. PGCs marked by SOX17 and/or Sda/GM2 are indicated by yellow dashed lines. Scale bar: 10µm. (C) 5hmC and 5mC levels determined by LC–MS/MS. Methylation levels are indicated relative to total levels of deoxyguanine (dG). P values are based on combined ANOVA and Holm's post hoc test.

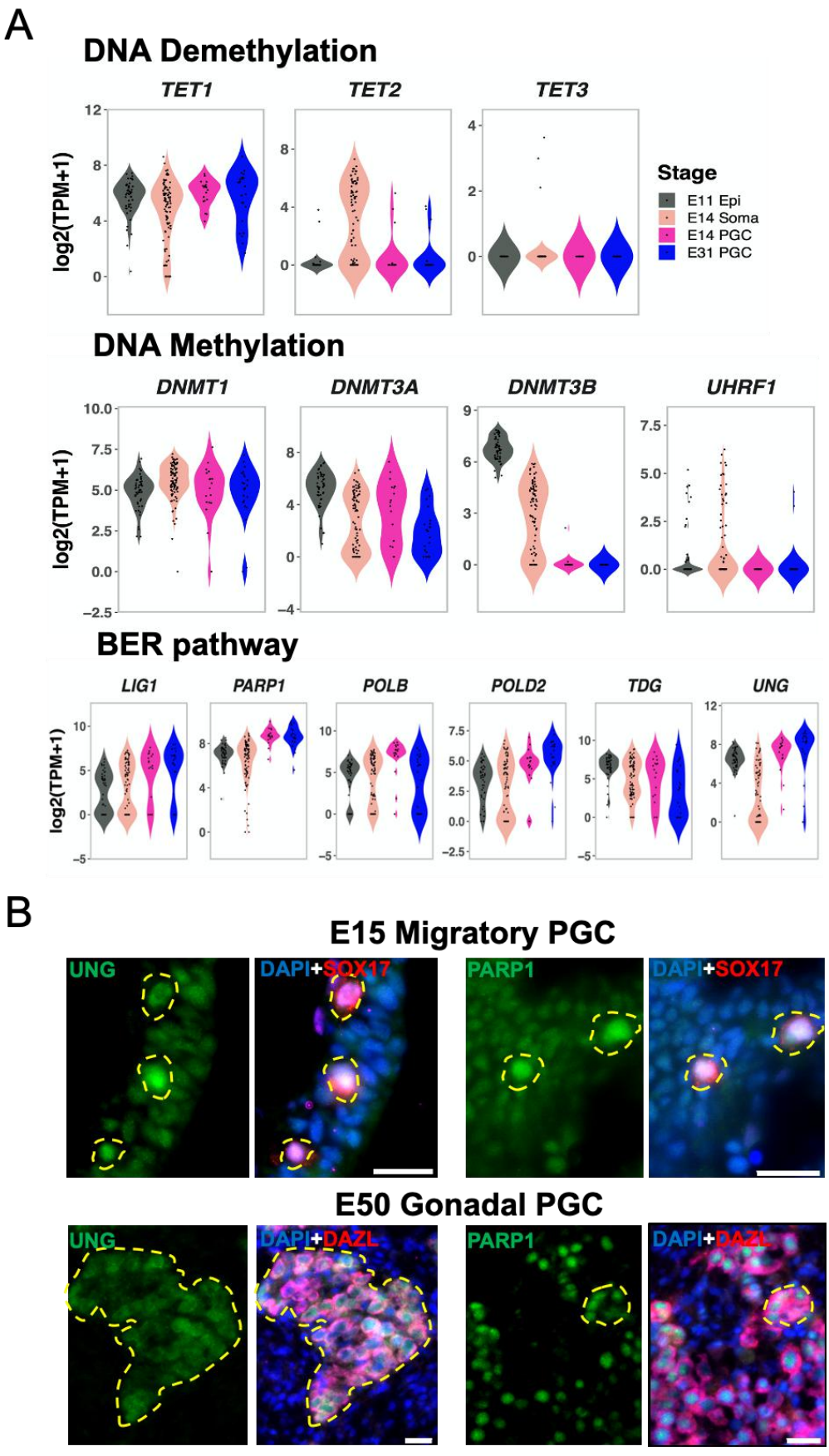

A
DNA Demethylation
TET1
TET2
TET3
log2(TPM+1)
Stage
E11 Epi
E14 Soma
E14 PGC
E31 PGC
DNA Methylation
DNMT1
DNMT3A
DNMT3B
UHRF1
BER pathway
LIG1
PARP1
POLB
POLD2
TDG
UNG
B
E15 Migratory PGC
UNG
DAPI+SOX17
PARP1
DAPI+SOX17
E50 Gonadal PGC
UNG
DAPI+DAZL
PARP1
DAPI+DAZL

Figure 2-34 (A) Expression of epigenetic modifiers for DNA methylation/demethylation and BER pathway components in E11 Epiblast, E14 Somatic Cells, E14 PGC and E31 PGC. (B) Immunofluorescence of BER pathway component UNG (left) and PARP1 (right) in E17 migratory and E50 Gonadal PGC. Yellow circle indicates PGC. Scale bar 20μm.

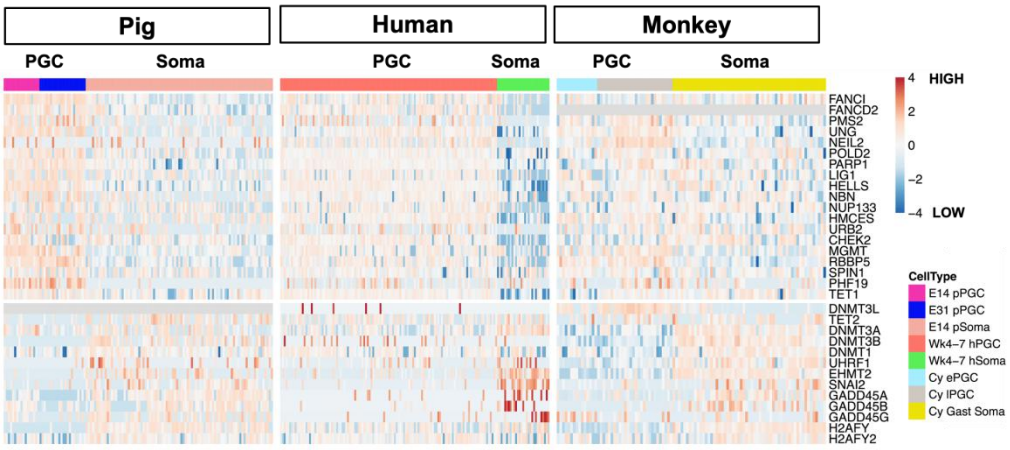


Figure 2-35 Expression heatmap of epigenetic modifiers and DNA repair genes that show differential expression in pPGC, hPGC and/or cyPGC compared to somatic tissue. "Cy ePGC": early cyPGC (E13-20); "Cy lPGC": late cyPGC (E36-55); "Cy Gast Soma": cy gastrulating cells (E13-20). Grey colour in heatmap indicates "NA", which indicates the gene is not annotated or not expressed in all cells examined in the species. As different expression units are used for the three species, values in the colour scale are replaced by HIGH and LOW.

2.3.2.2 Histone remodelling in early-migratory and gonadal pPGCs

Our lab previously demonstrated that the progressive reduction of H3K9me2 and its 'writer' G9a (*EHMT2*) starts in pPGC from E14 (Kobayashi et al., 2017)(Figure 2-35). Here, I further investigated the landscape of histone modification in pPGCs. Similar to hPGCs and mPGCs, H3K27me3 shows global enrichment in pPGC at the migratory stage (E17) (Figure 2-36 (A) and

(B)). At early gonadal stages (E25), H3K27me3 level starts to reduce and becomes almost absent in pPGC from E36 onwards (Figure 2-37 (A)), which is similar to hPGCs (Tang et al., 2015) but different from mPGCs and cyPGCs, as H3K27me3 is still detected in E13.5-15.5 mPGCs (Seki et al., 2005, Prokopuk et al., 2017) and E47-50 cyPGCs (Sasaki et al., 2016). Core components of PRC2 complex which catalyses H3K27me3 do not show differential expression in pPGCs (Figure 2-37 (B)). IF of EED, a core component essential for PRC2 assembly, indicates that this protein is present in most pPGCs at examined stages (from E15 to E50), which is different from previous report in mPGCs (Prokopuk et al., 2017). However, unlike in mPGCs (Prokopuk et al., 2017), no consistently differential staining can be observed between pPGCs and somatic cells (Figure 2-37 (D)). On the other hand, PHF19 and RBBP7, PRC2-associated cofactors that was suggested to be needed for PRC2 recruitment and activity, are enriched in pPGCs (Figure 2-37 (B)) (Ballaré et al., 2012, Brien et al., 2012, Li et al., 2017a, van Mierlo et al., 2019). DEG analysis also revealed dynamic changes of other histone and chromatin remodellers, such as upregulation of components of MII complex (*DPY30, RBBP5*) and SWI/SNF proteins *SMARCA5* and *HLTF* (Figure 2-38). Some of those genes, such as *RBBP5* and *PHF19*, also show upregulation in hPGC and cyPGC (Figure 2-35). *H2AFY*, on the other hand, is absent in epiblast and pPGC (Figure 2-36 (C) and Figure 2-37 (C)). MacroH2A1, the macro-histone variant encoded by *H2AFY*, is observed in gonadal mPGC (Nesterova et al., 2002, Hajkova et al., 2008), but absent in pPGC throughout all examined stages (Figure 2-37 (A)). Also, *H2AFY* is largely downregulated in early cyPGCs, although some of the late cyPGCs show mild expression of *H2AFY*. Interestingly, a homologues protein, macroH2A2, which is encoded by *H2AFY2*, is absent in gonadal hPGC (Tang et al., 2015). *H2AFY2* is mildly downregulated in pPGCs but not completely absent. Surprisingly, *H2AFY2* is not downregulated transcriptionally in gonadal hPGCs, suggesting the absence

of macroH2A2 is largely due to posttranscriptional regulation (Figure 2-35). Overall, these suggests that lack of macroH2A may be a feature of non-rodent PGC at early stages.

Overall, pPGC shows unique profiles of histone modifications compared to surrounding somatic cells, such as the upregulation and then absence of H3K27me3 during pPGC development. A novel epigenetic event, absence of macroH2A1 at both transcriptional and protein level, is also detected in pPGC of all stages. MacroH2A1 is suggested to be an epigenetic marker with various biological roles, including obstructing reactivation of pluripotent genes during reprogramming (Pasque et al., 2012, Gaspar-Maia et al., 2013, Pasque et al., 2011, Barrero et al., 2013a, Creppe et al., 2012), and impeding basal PARP1 activity which in turn affects DNA damage response and mitochondria respiration (Ladurner, 2003, Kustatscher et al., 2005, Hurtado-Bagès et al., 2018, Posavec Marjanović et al., 2017, Hodge et al., 2018). Absence of this marker, as well as other histone remodelling events, may create a unique histone/chromatin landscape for proper PGC development.

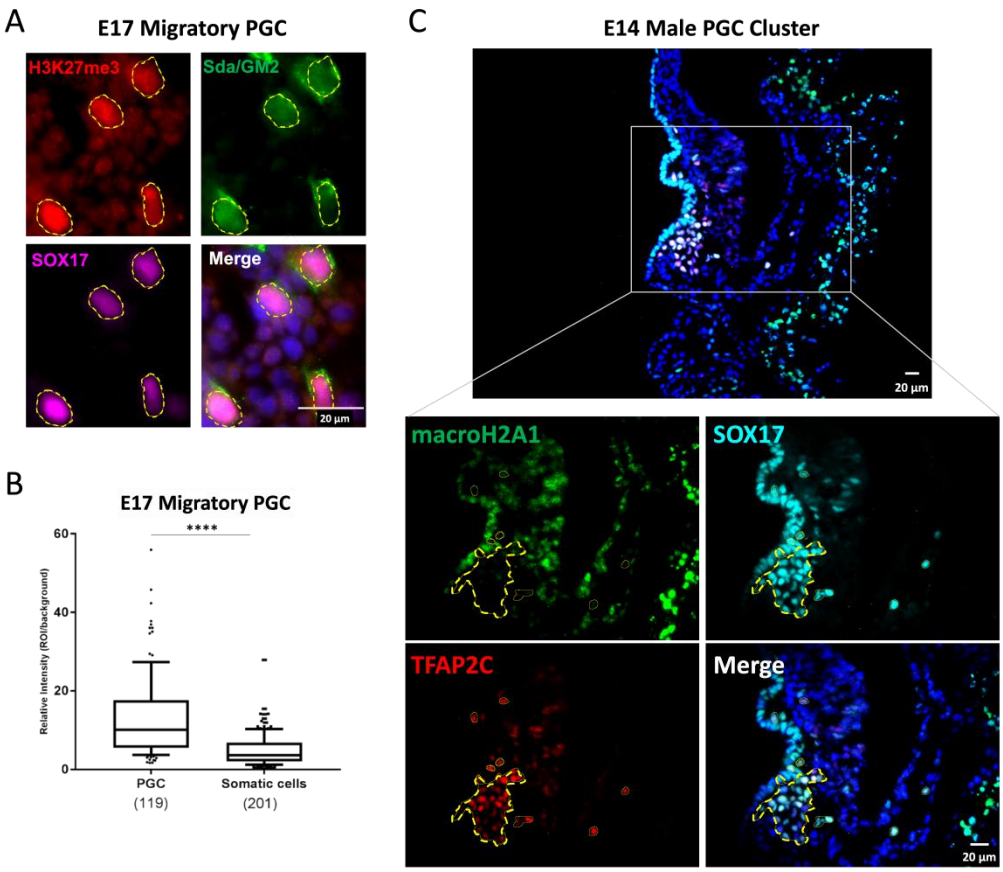


Figure 2-36 (A) and (B) show that immunofluorescence of H3K27me3 is enriched in early migratory pPGCs, while (C) macroH2A1, encoded by H2AFY, is absent in pPGCs from pre-migratory stage. pPGCs are marked by yellow dash lines. Scale bar: 20 µm

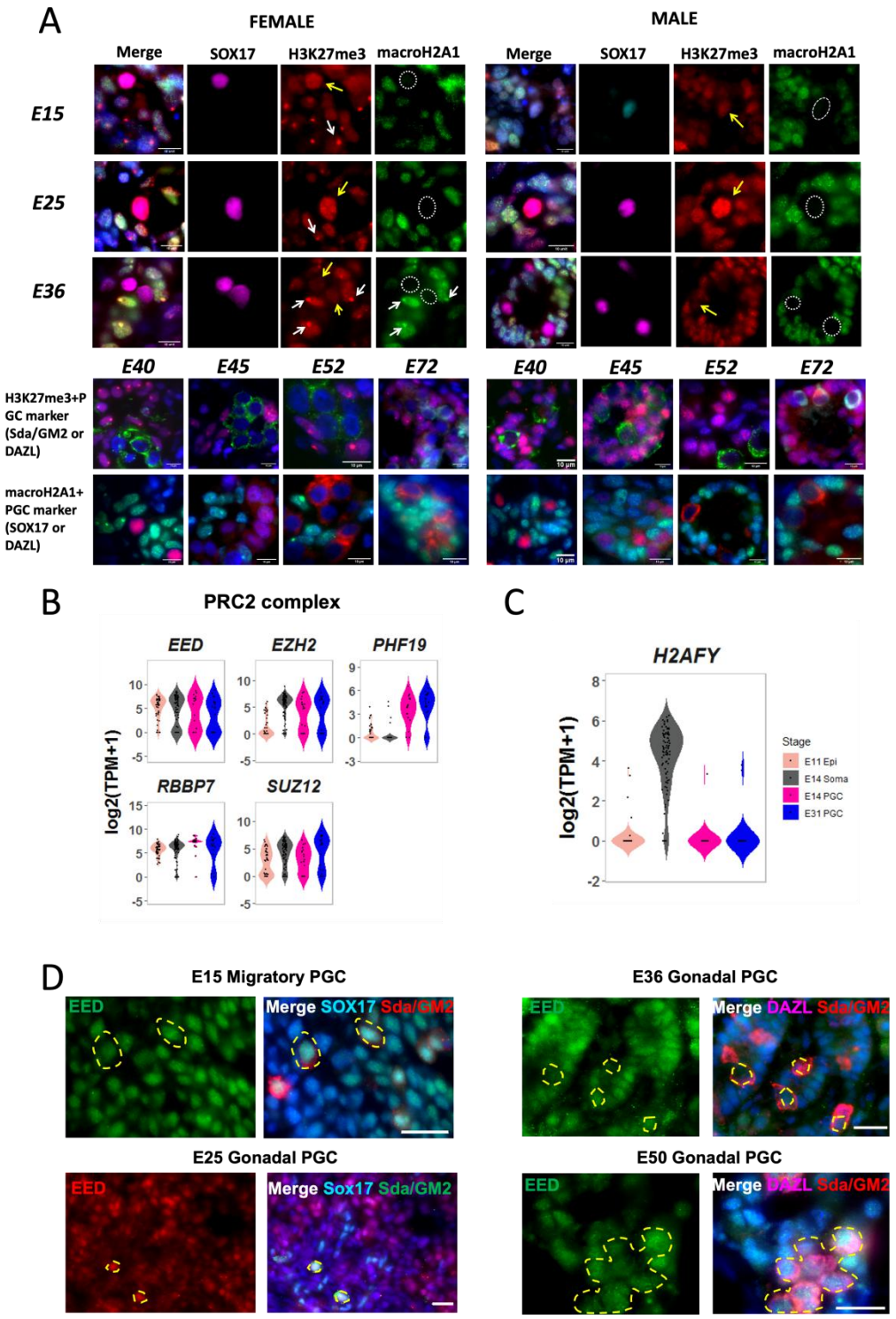


Figure 2-37 (A) H3K27me3 is enriched in migratory and early gonadal (E25) pPGCs, but absent in pPGCs from E36. In female pPGCs, the H3K27me3 spot which marks the Xi is largely lost from migratory stage. MacroH2A1 is absent in both male and female pPGCs in all examined stages. Yellow arrows and white dash circles indicate pPGC; White arrows indicate the H3K27me3 and macroH2A1 spot which marks the Xi. Scale bar: 10 μm (B) Expression profiles

of components of PRC2 complex. Key components EED, EZH2 and SUZ12 are present in pPGC but not differentially expressed compared to somatic cells. Cofactors PHF19 and RBBP7 are upregulated in pPGCs. (C) Consistent with the immunofluorescence result, H2AFY, which encode macroH2A1, is largely absent in pPGCs. (D) Immunofluorescence of EED in E15 migratory and E50 Gonadal PGCs. PGCs are marked by SOX17 and Sda/GM2 in E15 and E25, and by DAZL and Sda/GM2 in E36 and E50. Yellow circles indicate PGC. Scale bar: 20um.

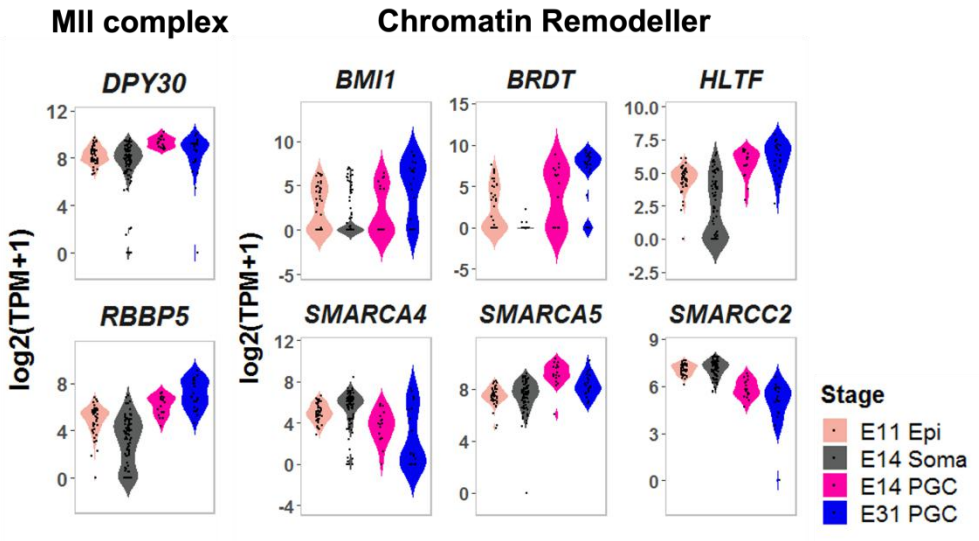


Figure 2-38 Expression profiles of other histone modifiers and chromatin remodellers in selected samples.

2.3.2.3 XCR starts in pre-migratory pPGCs

As we have previously showed that epigenetic reprogramming occurs in pre-migratory pPGC (Kobayashi et al., 2017), I investigated the transcriptional regulation and evidence of epigenetic reprogramming in pPGC. X chromosome reactivation (XCR) is a classic example of epigenetic reprogramming. In hPGCs, XCR and biallelic expression of X chromosome genes are detected from as early as 4 weeks (migratory stage) (Tang et al., 2015, Guo et al., 2015). In mouse, XCR is initiated in a small portion of migratory PGCs, but robust transcription (Sugimoto and Abe, 2007) and

translational activity (Chuva de Sousa Lopes et al., 2008) on the reactivating X chromosome (Xr) are only detected after reaching the genital ridge. X chromosome inactivation (XCI) is mediated via multiple layers of epigenetic repression, such as enrichment of H3K27me3 and DNA methylation. Inactive X chromosome (Xi)-associated H3K27me3 (H3K27me3 spot) is a commonly used marker for XCI, which is detected in over 90% of early-migratory (E7.5-7.75) mPGCs (Chuva de Sousa Lopes et al., 2008). IF of some pre-migratory pPGCs show no or faint H3K27me3 spot (Figure 2-39 (A)). Early-migratory pPGCs (E17) and early gonadal pPGC (E25) indicates that a significantly increased proportion (over 70%) of pPGC is absent of H3K27me3 spot (Figure 2-39 (B)). These suggests that XCR in pPGCs begin in pre-migratory stages and is evident in early-migratory stage. To further analyse XCR at transcriptional level in pPGCs, the sexual identity of E14 and E31 cells was inferred based on the cumulative level of Y chromosome genes per cell (Figure 2-39 (C))(Skinner et al., 2016). *XIST* expression is critical for XCI (Jonkers et al., 2008), which is repressed in majority of E14 and E31 female pPGCs, but not in female somatic cells, indicative of XCR occurring in pre-migratory and gonadal pPGC (Figure 2-39 (D)). Consistently, shown by female to male (F:M) ratio, X chromosome expression in E14 female PGCs is higher compared to male pPGCs, which further increases in E31 female PGCs, while the autosome expression is relatively close between male and female pPGCs and no gender differences are observed for either X chromosome or autosome expression in somatic cells (Figure 2-39 (E)). The ratio of X chromosome expression to total autosomal expression (X:allA) in female pPGCs is in general higher than female somatic cells and all male cells. At single cell level, the X:allA ratio is above 1 in all E31 female PGCs and some E14 female PGCs (Figure 2-39 (F)), which is also observed in female mPGCs and hPGCs at gonadal stages (Sangrithi et al., 2017). These further supports a specific X chromosome upregulation in female pPGCs from pre-migratory stage, while X chromosome expression in their male counterparts is relatively stable compared to somatic cells.

To rule out the possibility that the increased X:allA ratio and F:M ratio for X chromosome is due to expression changes of one active X chromosome, instead of biallelic expression from both X chromosomes in female PGCs, we further explored the X-chromosome gene expression at allelic resolution. E14 female somatic cells present a lower number of bi-allelically expressed single nucleotide polymorphisms (SNPs) (Figure 2-40 (A)), which is termed as 'ChrX escapers' in this project to distinguish them from the DNA methylation escapees in the next chapter. Studies show that 4-8% and 15-25% X-linked genes in mice and in human, respectively, escape XCI to some degree (Carrel and Willard, 2005, Cotton et al., 2013, Berletch et al., 2015, Balaton and Brown, 2016, Wainer-Katsir and Linial, 2016). ChrX escapers vary largely between tissues and species and have not been characterized in pig. Therefore, we categorized genes with biallelic SNPs in somatic pig cells as our X chromosome escapers. Consistent with aforementioned analysis, E14 female pPGCs have a large number of biallelic SNPs that are not ChrX escapers, which are similar to E31 female PGCs, indicating X chromosome reactivation at the transcriptional level in pPGC is evident at the pre-migratory stage (Figure 2-40 (A)). Biallelically expressed X-linked genes in female cells are further identified (Figure 2-40 (B)). All female pPGCs contain at least one biallelically expressed X-linked gene that is not found in somatic cells, i.e. non-ChrX escaper (Table 2-1). In contrast to a profound increase of biallelic expression in mouse gonadal PGCs compared to earlier stages (Sugimoto and Abe, 2007), the number of biallelically expressed, non-escaper X-linked genes is heterogeneous in pPGC and more similar between pre-migratory and gonadal stages, suggesting XCR may progress asynchronously and heterogeneously in pPGC.

Potential mechanisms of XCR are also investigated. In mice, most of the lncRNAs that are currently found to involve in XCI are located in X-chromosome inactivation centre (XIC) (Augui et al., 2011). During initiation of XCI, genes near the XIC in 3D space are first silenced, presumably due to the spread out of Xist from XIC (Engreitz et al., 2013, Pinheiro and Heard, 2017). There is no apparent relationship between the regional increase of F:M

ratio, which indicates X-linked gene reactivation, and the linear distance to the putative XIC in pig (i.e. homologues region to human XIC, highlighted in green) (Hwang et al., 2015) (Figure 2-41 (A)). Similarly, no apparent relationship between random XCI and linear distance to XIC was observed in pig peri-implantation embryos (Ramos-Ibeas et al., 2019). This suggests that the progression of XCR in pPGC is not correlated to the linear proximity to putative XIC. However, the possibility cannot be ruled out that the spread of XCR maybe correlated with the 3D distance to XIC in pig.

Recent studies have made large progress on revealing the Xist interactome via mass spectrometry-base techniques and gene screens (da Rocha and Heard, 2017). Xist interactors have important role in initiation and maintenance of XCI, and therefore their expression and activity may change dynamically during the XCR. However, common Xist interactors identified in both proteomic and genetic screens (i.e. High-confidence Xist interactors) do not show differential expression between PGCs and somatic cells (da Rocha and Heard, 2017)(Figure 2-41 (B)). A few Xist interactors show differential expression in all PGCs regardless of gender (Figure 2-41 (B)). Among these PGC-downregulated Xist interactors, there are SWI/SNF family members SMARCA4 (BRG1) and SMARCC2, as well as topoisomerase TOP2b. A study shows disruption of these proteins, combining with knockdown of Rad21, TOP1 and inhibition of DNA methylation machinery (Dnmt1), resulted in increased expression of a GFP reporter on the inactive X chromosome (Minajigi et al., 2015). As DNA methylation machinery is inhibited in PGCs due to absence of Uhrf1 (see section 2.3.2.1), it is possible that the downregulation of those Xist interactors, combined with inhibition of DNA methylation machinery, may contribute to the XCR in female PGCs.

Overall, both X chromosome expression and epigenetic status support a different XCR kinetics in pPGCs compared to mPGCs, of which XCR initiates rapidly in pre-/early migratory pPGCs and continues in gonadal pPGCs in a less stark manner. pPGC, combined with human in vitro systems, will be useful for understanding XCR in non-rodent mammals. XCR is part of the global epigenetic reprogramming in PGCs. This different XCR kinetics in

pPGCs further supports existence of key differences in the epigenetic reprogramming between mPGC and non-rodent PGCs. Furthermore, XCR itself is an excellent example of remodelling multiple epigenetic layers for transcriptional de-repression. The elucidation of this multi-layer regulation may also have great value in developing treatments for X-linked diseases.

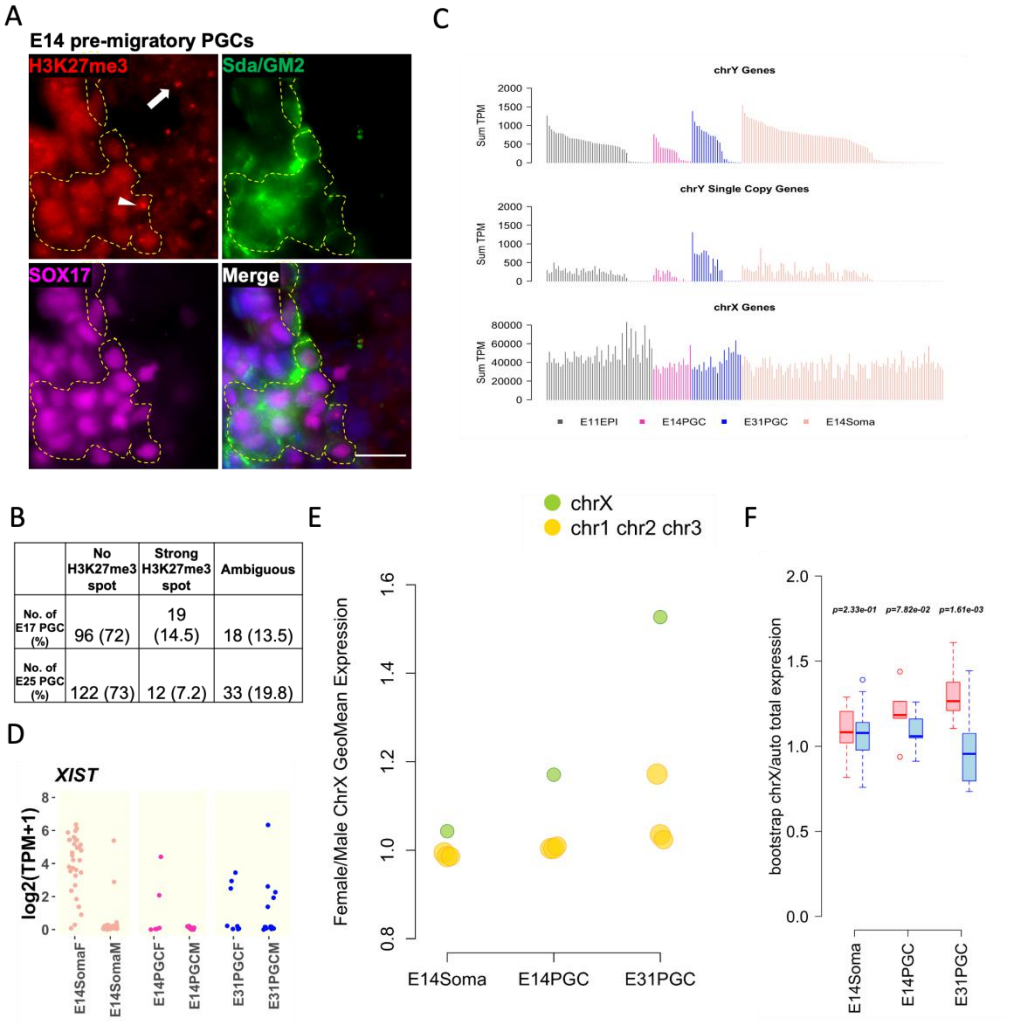


Figure 2-39 (A) Immunofluorescence shows that some pre-migratory pPGCs have no or weaker Xi-associated H3K27me3 spot and (B) a large portion of migratory and early gonadal pPGCs lack H3K27me3 spot Yellow circle indicates PGC. Xi-associated H3K27me3 are detected in somatic cells (arrow) and some pPGCs (arrowhead). Scale bar: 20µm. (C) A) Sum expression of (top) all Y-chromosome genes, (middle) single-copy Y-chromosome genes and (bottom) all X-chromosome genes. (D) Expression of XIST in E14 Somatic Cells, E14 PGCs and E31 PGCs. Xist expression is reduced in female pPGCs

from pre-migratory stage. (E) Female to male expression ratio of X chromosome genes vs. autosomes (chr1, chr2 and chr3) in E14 somatic cells, E14 PGCs and E31 PGCs. (F) Bootstrap of X:A ratio of E14 somatic cells, E14 PGCs and E31 PGCs. Each dot represents one cell. P values are determined by pairwise Wilcoxon Rank Sum Tests.

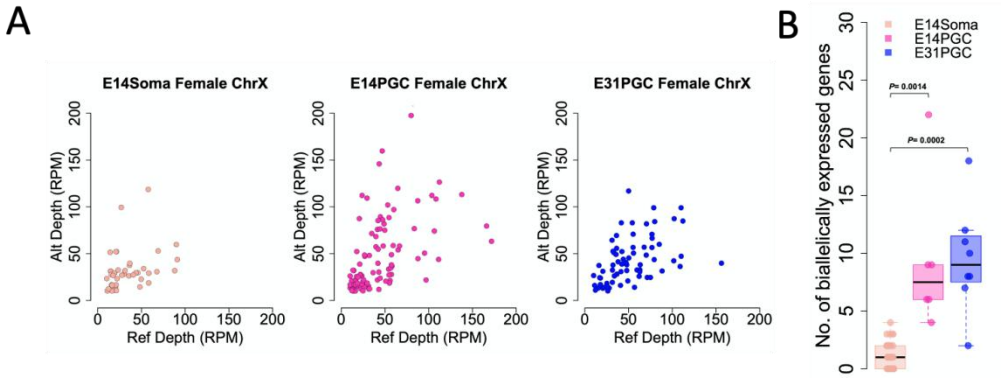


Figure 2-40 (A) Biallelically-detected SNPs on X chromosome identified in female E14 somatic cells, E14 PGCs, and E31 PGCs. Each dot represents one biallelically-detected SNP per cell. X axis: reads (RPM) that are mapped to the reference alleles; Y axis: reads (RPM) that are mapped to the alternative alleles. Escapers are removed in the PGC plots. (B) Number of biallelically expressed genes in E14 PGCs, E31 PGCs and E14 Somatic cells. P value is determined by Kruskal-Wallis test followed by Dunn's test.

Table 2-1 Escapers and biallelically expressed genes in E14 and E31 pPGCs. '-/-': gene not expressed; '+/-': gene mono-allelically expressed; '+/+': gene biallelically expressed. One E14 pPGC (AJ17) shows larger number of biallelically expressed genes.

	E14 PGC						E31 PGC							
	AJ17	AK26	AJ31	AK82	AJ110	AJ24	GF17	GF15	GF12	GF23	GF13	GF24	GF14	GF26
Escaper														
AP1S2	-/-	-/-	+/-	-/-	+/-	-/-	+/-	+/-	-/-	+/-	-/-	-/-	-/-	-/-
DDX3X	+/-	+/-	+/+	+/-	+/-	+/+	+/-	+/-	+/-	+/+	+/-	+/-	+/+	+/-
FUNDC1	+/+	+/+	+/+	+/+	+/+	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
MECP2	-/-	-/-	+/-	+/-	+/-	+/-	+/-	-/-	-/-	-/-	+/-	+/-	-/-	+/-
SSR4	+/+	+/-	+/-	+/+	+/-	+/-	+/+	+/+	-/-	+/+	-/-	+/+	+/+	+/+
ENSSSCG00000030283	+/-	+/-	+/-	+/-	+/+	+/-	-/-	+/-	-/-	+/-	+/-	+/-	+/-	-/-
BCAP31	+/+	+/-	+/-	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
RPS4X	+/-	+/-	+/+	+/-	+/-	+/-	+/+	+/+	+/+	+/+	+/-	+/+	+/+	+/+
UBA1	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/-	+/-	+/-	-/-	+/+
ENSSSCG00000038963	-/-	-/-	+/-	+/-	+/-	-/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
USP9X	+/-	+/-	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
RPL10	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
Common														
CLCN4	+/+	+/-	+/-	+/-	+/+	+/-	-/-	+/-	+/-	+/-	-/-	-/-	+/-	+/+
MID1	+/+	+/-	-/-	-/-	+/-	+/-	-/-	+/-	-/-	+/+	-/-	+/-	-/-	+/-
NDUFB11	+/+	+/-	+/-	+/-	+/-	+/-	-/-	+/+	+/-	+/+	-/-	-/-	+/+	+/+
CXorf56	+/+	+/-	+/-	+/+	+/+	+/-	+/-	+/-	+/+	+/-	+/-	+/-	+/+	+/-
LAMP2	+/+	+/-	+/-	+/+	+/-	+/-	+/+	+/+	+/-	+/+	+/-	+/-	+/+	+/-
SRPK3	+/+	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/+	+/+	+/+	+/+	+/-
IDH3G	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/-	+/-	+/+	+/-
SASH3	+/-	+/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/+	+/-	+/-	+/+	+/+
ZDHHC9	+/-	+/-	+/+	-/-	+/-	+/-	-/-	-/-	+/-	+/+	-/-	-/-	+/-	+/-
ATP6AP2	+/-	+/-	+/-	+/+	+/+	+/+	+/-	+/+	+/-	+/+	+/-	+/+	+/-	+/+
E14 PGC-specific														
RBBP7	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-	+/-
PDK3	+/+	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
KDM6A	+/+	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-	-/-	+/-	-/-	-/-
PRPS1	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
PSMD10	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-	+/-
AMOT	+/+	+/+	+/-	+/-	+/-	+/-	-/-	-/-	+/-	+/-	+/-	+/-	+/-	-/-
ENSSSCG00000038477	+/+	+/-	-/-	+/-	-/-	-/-	-/-	-/-	-/-	+/-	+/-	+/-	+/-	+/-
ENSSSCG00000012731	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-	+/-
ENSSSCG00000012746	+/+	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
GDI1	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
TSR2	+/-	+/-	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-
ENSSSCG00000012654	+/-	+/-	+/+	+/-	+/-	+/-	-/-	-/-	+/-	-/-	-/-	-/-	-/-	+/-
TSPAN6	+/-	+/-	+/-	+/+	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-
FTSJ1	+/-	-/-	-/-	+/-	+/-	+/+	-/-	+/-	-/-	-/-	-/-	-/-	-/-	-/-
E31 PGC-specific														
TSPAN7	-/-	+/-	-/-	-/-	-/-	-/-	+/+	-/-	+/-	+/-	-/-	-/-	-/-	-/-
PGK1	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/-	-/-	+/+	+/-	-/-	+/+	+/+
ALG13	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/-	+/-	+/+	+/-	+/-	+/-	+/-
PRPS2	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/+	-/-	+/-	-/-	-/-	-/-	-/-
PAGE2B	-/-	-/-	-/-	-/-	-/-	-/-	-/-	+/+	-/-	-/-	-/-	+/-	-/-	+/-
PLS3	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/+	+/+	+/-	+/+	+/-	+/-
ENSSSCG00000012757	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/-	+/+	+/-	-/-	+/-	+/-
ENSSSCG00000012119	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/-	+/-	-/-	+/-	+/-
TRAPPC2	+/-	+/-	-/-	+/-	+/-	+/-	+/-	+/-	+/-	+/+	+/-	-/-	+/-	+/-
USP26	-/-	-/-	-/-	+/-	-/-	-/-	-/-	+/-	-/-	+/+	-/-	+/+	+/-	+/-
CDKL5	-/-	+/-	-/-	+/-	-/-	-/-	+/-	+/-	-/-	+/-	+/-	+/+	+/-	+/-
ENSSSCG00000035908	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-/-	+/-	+/+	+/-
CUL4B	+/-	+/-	-/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+/+

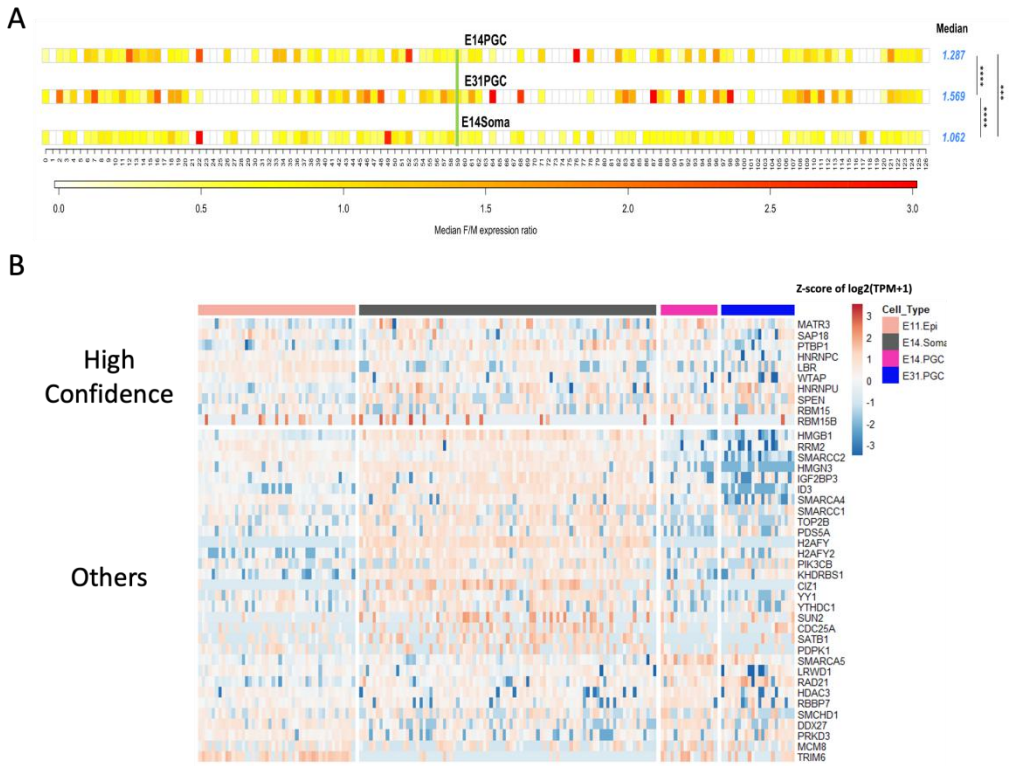


*Figure 2-41 (A) Median female to male expression ratio across X chromosome in E14 PGC, E31 PGC and E14 Somatic cells. p value. *** p < 0.001; **** p < 0.0001 by pairwise wilcoxon test. Green line indicates estimated position of XIST inactivation. Activation of ChrX genes do not show clear proximity to the XIST loci. (B) Expression of reported Xist interactors in indicated samples. The high confidence interactors do not show differential expression in pPGCs, but the downregulation of other Xist interactors such as TOP2B and SMARCA4, may contribute to XCR.*

2.4 Summary

In this chapter, by analysing the high-resolution scRNA-seq data, I explored the transcriptional network and epigenetic reprogramming in pre-/early migratory and gonadal pPGCs. Unique events and factors during pPGC development are identified, many of which are also observed during hPGC and/or cyPGC development, suggesting the pig is a relevant system for studying specification and development of non-rodent germline.

Transcriptional network of pre-/early migratory and gonadal pig PGCs

As expected, pPGCs are specified at the posterior end of a flat-disc embryo, surrounded by potentially gastrulating primitive streak and amnion precursors. PGCs show large transcriptional difference compared to these surrounding cell lineages at pre-migratory stage. Both pre-migratory and gonadal pPGCs express known germline markers including *SOX17*, *BLIMP1*, *TFAP2C* and *NANOS3* and lack *SOX2*, which is similar to primate PGCs but different to mPGCs. UMAP analysis also support the similarity between pPGCs and primate PGCs as pre-migratory and gonadal pPGCs cluster nicely with cyPGCs of similar stages and separate from pig and Cy monkey somatic cells. Nevertheless, pPGCs still have some unique features. For example, *PRDM14* is largely absent in pPGC, which suggests different regulatory mechanisms for epigenetic reprogramming compared to in mPGCs (Yamaji et al., 2008) (Figure 2-42). Although pPGCs exhibit some pluripotency features similar to E6 ICM and E8 Epiblast, they do not show the high expression of naïve pluripotency markers *KLF4* and *TFCP2L1* which was observed in both hPGCs and cyPGCs. It is worth noting that those pluripotency markers or features may not be essential for PGC specification or maintenance of naïve pluripotency, as recently reported (Hancock et al., 2020). Multiple genes shared by pPGCs and pluripotent cells are related with retrotransposon repression. It is possible that the shared genes between pPGCs and pluripotent cells reflect a similar epigenetic landscape, including large-scale DNA demethylation and X chromosome reactivation. Unique and stage-specific surface markers such as *GPR50* are also identified, which may be useful for future isolation of migratory pPGC.

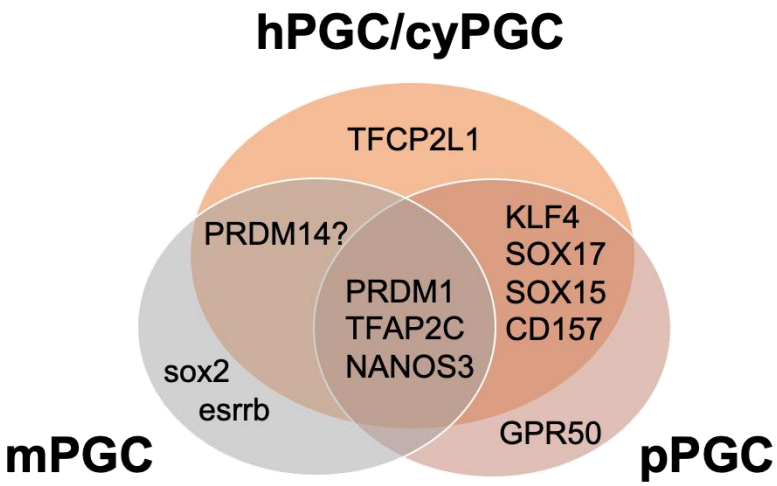


Figure 2-42 Schematic of unique or common genes between pig, human/Cy monkey and mouse PGC.

The transcriptional difference between pPGCs and surrounding cell lineages reveal multiple PGC-specific events, including metabolic and epigenetic reprogramming. From pre-migratory stage, pPGCs exhibit evident downregulation of glycolysis and cholesterol synthesis, which is also observed in cyPGCs. Mild upregulation of some components of TCA cycle and ETC is also observed in pPGC, however most of the components in these two pathways are not differentially expressed. Surprisingly, mtDNA-encoded protein-coding genes show downregulation in E14 but not E31 pPGC, which may reflect the stringent mtDNA bottleneck observed in hPGCLCs (Floros et al., 2018). Further investigations will be required to validate those metabolic changes at the metabolite and protein levels.

Epigenetic reprogramming in pPGCs was investigated in three aspects, DNA demethylation, histone remodelling and XCR. In general, pPGCs exhibit unique tempo and mechanisms of epigenetic reprogramming compared to mPGCs, including active DNA demethylation and evident onset of XCR at pre-migratory stage, enrichment followed by absence of H3K27me3 from gonadal stage and persistent lack of macroH2A1. Potentially acting downstream of active DNA demethylation and protecting genome integrity of

germline during major epigenetic changes, BER pathway and other DNA repair components are also upregulated from pre-migratory stage. The regulation and mechanisms of the early onset of XCR in pPGC, such as the dynamics of lncRNAs and their interactors, may also give some insight on X chromosome dynamics from an evolutionary perspective as XCR may be regulated differently among species like previously reported for XCI (Vallot et al., 2017, Okamoto et al., 2011). Some of those unique epigenetic features are also observed in hPGCs and/or cyPGCs. It is worth noting that there appears to be some discrepancies between hPGCs and cyPGCs, including the highly and consistently expressed DNMT3L in early and late cyPGCs as well as moderate H3K27me3 level in E47-50 cyPGCs (Sasaki et al., 2016), both of which were not observed in hPGCs.

Overall, the high-resolution scRNA-seq data of pPGC provides a useful resource for investigating non-rodent PGC development. The unique events identified during pPGC specification and early development are worth being further examined in early cyPGCs and during hPGCLC induction, which will extend current knowledge of mammalian PGC development and provide insights for hPGC development.

Chapter 3 DNA Methylome of Gonadal pPGC

3.1 Introduction

Epigenetic reprogramming is a key event in mammalian PGC, characterized with global DNA demethylation, histone remodelling and XCR. This reprogramming is suggested to prevent the transmission of potential deleterious parental epigenetic information to the next generation. However, multiple studies in mammals reported inherited phenotypic variations, which are induced by environmental factors and cannot be explained by genetic variation (Carone et al., 2010, Braunschweig et al., 2012, Feeney et al., 2014, Chen et al., 2016c, Huypens et al., 2016, Bohacek et al., 2016). Furthermore, recent studies in human and mouse have identified regions that retain DNA methylation during global demethylation in PGCs (Tang et al., 2015, Gkountela et al., 2015, Guo et al., 2015, Kobayashi et al., 2013). Those regions are termed 'escapees' and have the potential to be carriers of epigenetic information across generations. These escapees show some conservation between human and mouse at orthologous gene loci but not at sequence level. Escapee genes in human are shown enriched for genes related to metabolic and mental disorder (Tang et al., 2015). However, because of the key differences reported between mouse and hPGC development (Tang et al., 2015, Irie et al., 2015, von Meyenn et al., 2016), the anatomical and physiological dissimilarities between mouse and human, as well as the poor conservation of escapees at sequence level, mouse may not be the ideal model for further exploration of the escapees in transgenerational inheritance experiments. On the other hand, early PGC development in pig has shown striking similarity with human and monkey in vitro induction (Kobayashi et al., 2017), and pig has been used as large animal model for obesity and diabetes mellitus (Robert et al., 2015, Walters et al., 2011, Bassols et al., 2014, Schachtschneider et al., 2015). Therefore, the pig is a relevant model for studying potential roles of escapees in disease inheritance in humans. To investigate global DNA demethylation and identify potential escapee in pig, I obtained DNA methylome of E35 gonadal pPGCs and somatic cells using PBAT-seq. E35

pPGCs are of equivalent stage to E13.5 mPGCs and Week 7 hPGCs, of which DNA demethylation is complete, and therefore contain the lowest methylation level (Hyldig et al., 2011a) that are optimal for identification of the escapees.
The aim of this chapter is to characterize gonadal pPGC methylome and to identify escapees in pPGCs. The methylome and escapees of gonadal pPGCs will be further used to compare with those of hPGC and mPGC to identify the similarities and differences between species. The methylome data will also be correlated with transcriptome data of pPGCs to reveal roles of DNA methylation for transcriptional regulation.

3.2 Material and Method

3.2.1 **Gonad collection and digestion and FACS for high purity pPGCs**

All the procedures involving animals have been approved by the School of Biosciences Ethics Review Committee, The University of Nottingham. The gonad collection and digestion as well as FACS procedures for single cell samples (section 2.2.1 and 2.2.3) were also used for E35 PBAT samples with a few modifications. Two crossbred Large White and Landrace sows were inseminated and sacrificed 35 days after the insemination. As one sow was not pregnant, all E35 embryos were collected from one litter and dissected for gonads. Tails of E35 embryos were used for sexing the embryo via PCR with AMELX primers (Sembon et al., 2008). Hereafter, pairs of E35 pig gonad from individual embryos were processed separately. E35 gonads were stored in foetal bovine serum (FBS) at 4°C overnight before digestion on the next day. For gonad digestion, Tissues are chopped into small pieces with scalpel in collagenase IV (2mg/ml in DMEM), and kept at 37°C for 30-40 min with vortexing every 7-8 min. Tissue is further digested by Tryple for 3 min at 37°C. After removal of Tryple, the cell suspension was filtered through 40-µm Pre-Separation Filters to generate single-cell suspension. Single-cell suspension were incubated with pPGC marker Sda/GM2 (1:50) for 30 min at 4°C and then Donkey anti-Mouse IgG (H+L) Alexa Fluor® 488 conjugate (Life technologies, A-21202) for 20 min on ice. MoFlo™ XDP cell sorter is used for FACS. Stringent gating and double-sorting ($1^{\#}$ enrichment + $2^{\#}$ purity mode)

are applied to ensure sample purity without major compromise of the quantity of the sample. After second sort, round 200 Sda/GM2^{+} cells are collected on slide for manual check of purity.

3.2.2 PBAT Library Preparation and sequencing

To generate methylome data comparable to data of hPGCs and mPGCs, PBAT libraries are prepared as previously described (Tang et al., 2015) with a few modifications (Clark et al., 2017) (Figure 3-3). In brief, FACS-sorted Sda/GM2^{+} cells (PGC) and Sda/GM2^{-} cells (somatic cells) were incubated with lysis buffer (0.1% SDS, 50 ng/ml carrier RNA (QIAGEN) and 1 mg/ml proteinase K (Zymo Research) in DNase-free water for 60 min at 37°C. Around 1% of unmethylated lambda phage DNA (0.1-0.2 ng) (Promega) was spiked into the sample before bisulfite treatment with the Methylcode Bisulfite Conversion Kit (Invitrogen). Manufacturer's instructions were followed for the bisulfite treatment, except for the incubation time, which were extended to 3.5 hours as described previously (Tang et al., 2015) to facilitate fragmentation and minimize false detection of methylation.

Bisulfite-treated DNA was re-annealed to double-stranded DNA using Klenow fragments (3'–5' exo-) (New England Biolabs) with a 5' biotin tagged primer consisted of an Illumina adaptor followed by 6 random nucleotides at the 3' end (Bio_N6: 5'-biotin-CTACACGACGCTCTTCCGATCTNNNNNNN-3'). The biotinylated first strand molecules were captured using Dynabeads M280 Streptavidin (Invitrogen) and then reannealed to double-stranded DNA again using Klenow fragments (3'–5' exo-) with random primers containing Illumina adaptors (Rev_N6_PE: 5'-TGCTGAACCGCTCTTCCGATCTNNNNNN-3'). Template DNA strands were then extended and amplified for 11 PCR cycles using KAPA HiFi HotStart kit (Kapa Biosystems) with the Illumina primer PE 1.0 (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATC*T-3') and customized primer iPCRTag (5'-CAAGCAGAAGACGGCATACGAGATAACGTGATGAGATCGGTCTCG GCATTCCTGCTGAACCGCTCTTCCGATC*T-3') (Clark et al., 2017). Agencourt AMPure XP beads (Beckman Coulter Genomics) are used for size

selection of the library. Size distribution of PBAT libraries were examined by Agilent High Sensitivity DNA Kit (Agilent). Concentrations of PBAT libraries were quantified by qPCR using NEBNext Library Quant kit (New England Biolads). Libraries, with 12% PhiX spike-in (Illumina), were subjected to pair-read 150 bp sequencing on HiSeq 4000 sequencing system (Illumina). Sequencing statistics and bisulfite conversion rate are summarized in Appendix Table 2.

3.2.3 **Bioinformatic analysis**

The quality of raw reads was determined by FastQC to ensure that the experimental setup and sequencing were successful. Raw reads were trimmed by skewer first to remove adaptor sequences and reads with low sequencing qualities (Jiang et al., 2014). Then, both the ends of paired-end reads were trimmed to improve the mapping efficiency. Forward reads were trimmed by 10 bases at the beginning, while reverse reads were trimmed by 5 bases at the end.

Trimmed reads were directionally aligned against the pig genome (Sus Scrofa v11.1) in the paired-end mode by hisat2 using Bismark pipeline with *--pbat*. --score_min was L,0,-0.4. *deduplicate_bismark* was applied to remove the potential PCR duplicates with default settings (Krueger and Andrews, 2011). Unmapped reads were re-aligned with the same parameters in the single-end non-direction mode to rescue misaligned paired-end reads due to the incorrect insert size resulting from the narrow sequencing area. The single-end alignment was merged with the paired-end alignment after deduplication.

To compare the pig PBAT datasets with those from human (Tang et al., 2015) and mouse (Kobayashi et al., 2013), reads were trimmed up to 100 nt for all three species and were mapped via single-end only and sampled to the same depth.

The detection of methylated cytosines was done by *bismark_methylation_extractor*, which can provide the genome-wide cytosine methylation status. The spike-in unmethylated lambda phage DNA was also

included in the analysis to examine the efficiency of bisulphite conversion in the samples. In the following analysis, only methylated cytosines at CpG context are used. Methylated cytosines at non-CpG context are very rare in somatic cells and almost absent in pPGCs, and therefore not further analysed.

The methylation level was calculated by the module *roimethstat* of *MethPipe* according to the locations of CpG islands and CGI shores, the genomic features and by the repeat density (Song et al., 2013). Annotations of CpG islands, genes, promoters and repeat regions were downloaded from UCSC and Ensembl databases. Promoter regions were defined as sequences located between 1,000 bp upstream and 500 bp downstream of a transcription start site. Promoters with high-CpG content (HCP) contain a 500 bp region with a CpG ratio larger than 0.75 and a GC content larger than 55 %. Promoters with low-CpG content (LCP) do not contain a 500-bp region with a CpG ratio larger than 0.48. Intermediate-CpG promoters (ICPs) are neither HCP nor LCP. Methylation level of previously reported differentially methylated regions (DMRs) (Congras et al., 2014, Hyldig et al., 2011a, Canovas et al., 2017, Yu et al., 2018a) in pig imprinted genes (Morison et al., 2005) and *XIST* gene (Zou et al., 2019) were also calculated. If DMRs have not been identified in pig imprinted genes, homologous regions inferred human or mouse DMRs were used (Tomizawa et al., 2011, Okae et al., 2014, Court et al., 2014). Methylation-sensitive genes that have also been reported previously in hPGCs (Tang et al., 2015) and mPGCs (Hill et al., 2018) were also examined. CGI-containing X chromosome promoters were also examined, among those promoters the partially methylated (30%–70% methylation) ones in both male and female gonadal somatic cells are plotted in heatmaps.

Significantly hypermethylated regions (HyperMR) were identified by the *hmr* function of *MethPipe*. Escapees were defined as HyperMRs which have more than 20% of CpGs with >= 5x and with at least 30% methylation level in human and 15% in pig and mouse. TE-poor escapees were defined as less than 10% of regions overlapped with repeats. TE-rich escapees were defined as at least 10% of regions overlapped with repeats.

To examine the methylation level of pig TE-poor escapees in early blastocyst, processed PBAT data of three in vivo E7.5 blastocyst was downloaded. Genome-wide cytosine methylation status of the blastocysts were extracted by Bismark as mentioned above. As the processed blastocyst data has been aligned to susScr3 (i.e. Sus Scrofa v10.2) assembly not Sus Scrofa v10.2, TE-poor escapee regions were lift over to susScr3 using UCSC *liftover* tool.

For TE expression, repeat regions were downloaded from UCSC database including all the sub families. *featureCounts* (Liao et al., 2014) was used to determine the number of reads aligned to each region with *-M* option. Expression level was calculated and normalised by Reads Per Kilobase Million (RPM).

3.3 Results

3.3.1 Purity and yield of FACS-sorted Sda/GM2$^+$ cells (pPGC)

The purity of PGC samples is critical for generating high-quality PGC methylome. As gonadal PGCs are extremely demethylated (<10% methylation) while somatic cells are highly methylated (~80% methylation) (Tang et al., 2015), minor contamination of somatic cells in PGC sample can lead to large shift in the methylation data of PGCs. FACS-sorting with double PGC surface markers, as described by Tang, et al., would be ideal for yielding pure PGC samples. However, Sda/GM2 is currently the best surface marker available for pPGCs (Klisch et al., 2011). Other surface markers, such as SSEA-1 (Manku and Culty, 2015, Kim et al., 2013c), are not detected in over 80% of gonadal pPGCs, and therefore will lead to low cell yield if used for FACS-sorting (Figure 3-1). To ensure the purity of pPGC samples, double FACS-sorting was used. The first sort with enrichment mode maximizes the cell yield, while the second sort with purity mode will further purify the Sda/GM2$^+$ samples by excluding ambiguous droplet (such as droplets containing one positive cell and one negative cell) (Figure 3-2 (A)). Manual check of the double-sorted Sda/GM2$^+$ cells indicates average of 98.5% purity (Figure 3-2 (B)). Around 1000-7000 Sda/GM2$^+$ and 5000 Sda/GM2- cells were collected from

individual embryos as our PGC and somatic samples for PBAT library, respectively (Appendix Table 3).

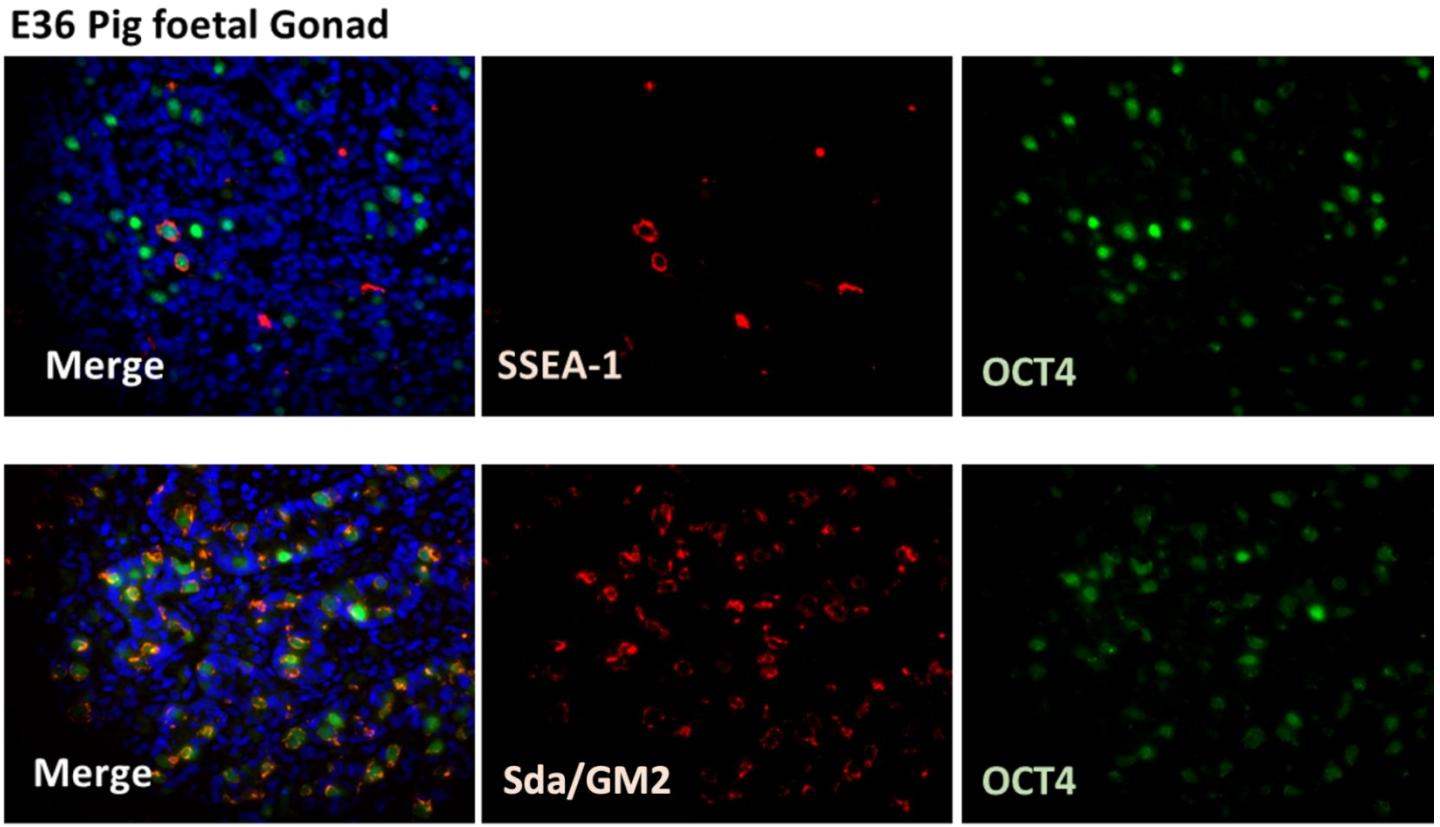


Figure 3-1 Immunostaining of E36 gonadal pPGCs with surface marker Sda/GM2, SSEA-1 and nucleus marker OCT4. Surface marker CT1 shows large overlap with nucleus marker OCT4, whereas SSEA-1 is only detected in a few OCT4-positive cells

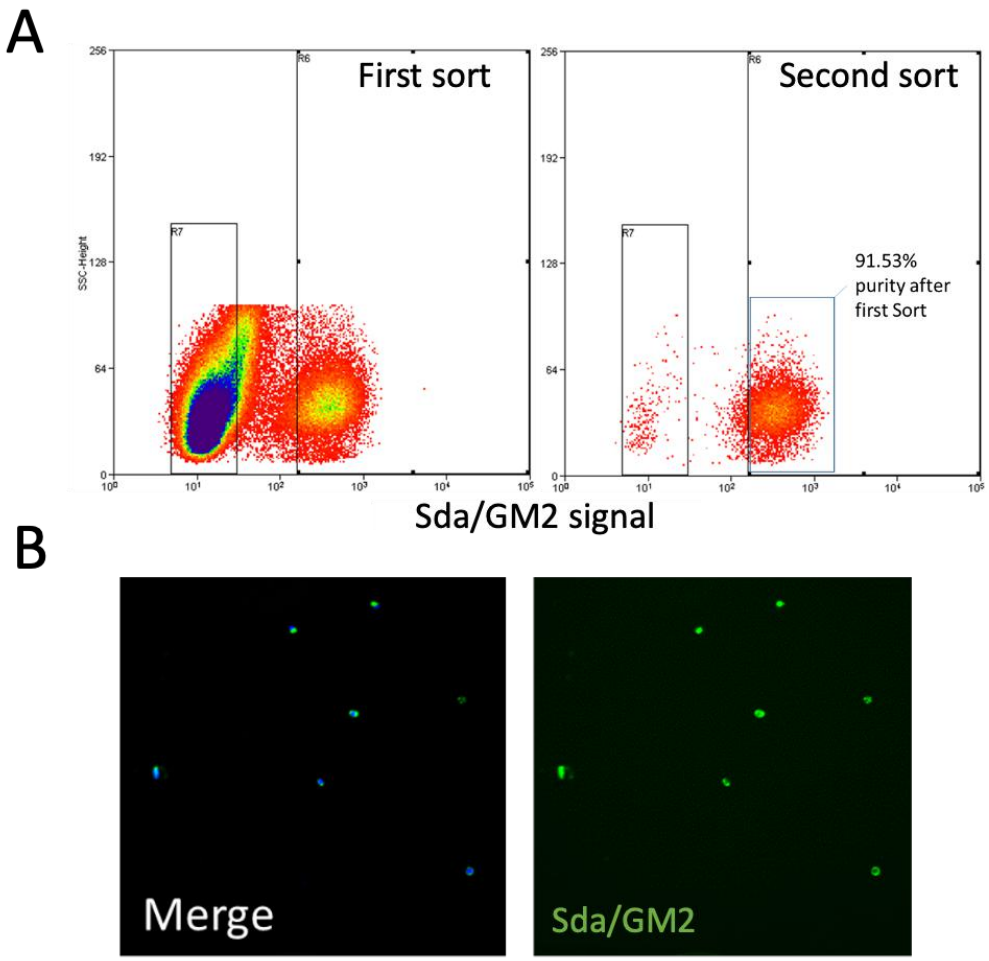


Figure 3-2 Purity of FACS-sorted PGC samples.(A) is exemplary first and second FACS sort of one sample, respectively. A small percentage of Sda/GM2$^-$ cells are still present after the first sort (91.53% are Sda/GM2$^+$), largely due to the enrichment mode, and therefore will be removed in the second sort with purify mode. (B) is an exemplary image of FASC-sorted Sda/GM2$^+$ cells on slide.

3.3.2 Modification of PBAT library Preparation and Quality of the PBAT-seq data

PBAT libraries were prepared as previously described (Tang et al., 2015) with a few modifications. I used primers containing 6 random nucleotides, instead of 4 or 9, to balance between library yield and sequence that needs to be trimmed per read before mapping (Clark et al., 2017) (Figure 3-3). Template extension and PCR amplification are also combined, as described by Clark et al. On average, 74% fragments in the PBAT libraries are within 200-600bp (Figure 3-4 (A)) and average size-corrected library concentration is 158.5 nM. These indicates the PBAT libraries are suitable for sequencing on Hiseq 4000. Compared to the size distribution shown by Clark et al., the size distributions of our PBAT libraries are slightly right shifted, indicating larger proportion of longer fragment. However, the size distribution profiles can only be slightly changed by time of bisulfite incubation or concentration of starting material (Figure 3-4 (B)). Different size selection methods using AMPure beads only remove the longer fragment but not increase the quantity of fragment of appropriate size, and therefore does not largely improve the library quality. Considering that extended bisulfite incubation or reduced starting material may increase inappropriate conversion of 5mC to T (i.e. false negative) (Holmes et al., 2014), and the size distribution of currently generated PBAT libraries are acceptable, we continued to sequence those libraries.

In total, six PBAT libraries are sequenced, including two male and two female Wk5 (E35) pPGCs isolated from single embryos as well as one male and one female Wk5 somatic cells (Appendix Table 3). The sequencing statistics shows high CpG coverage (over 60% CpG sites are covered by 5x) and appropriate bisulfite conversion rate (99.1% on average), which are comparable to previous hPGC (Tang et al., 2015, Guo et al., 2015) and mPGC methylome (Kobayashi et al., 2013) and indicate appropriate mapping and high-quality data (Appendix Table 2).

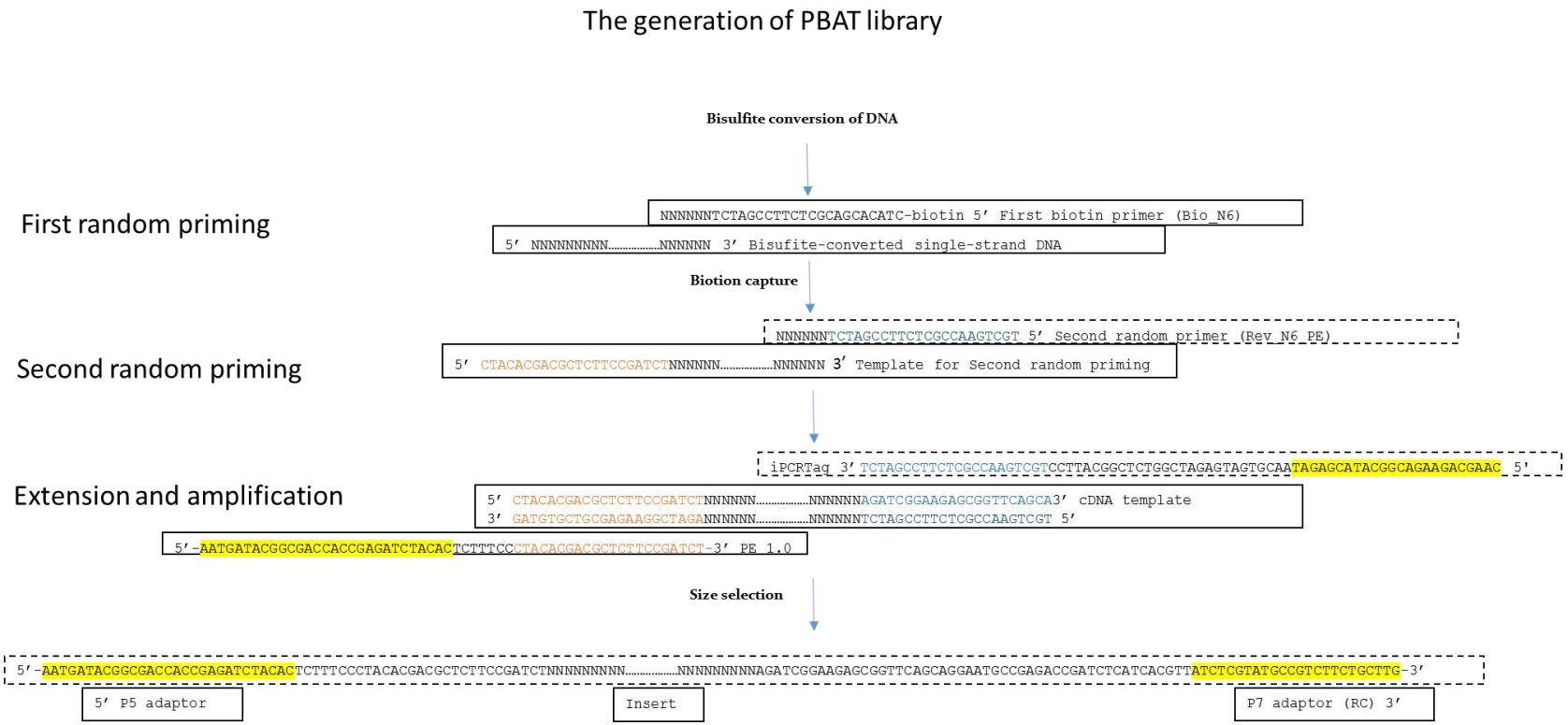


Figure 3-3 Brief illustration of PBAT library Generation. During generation of PBAT library, first and second random priming with attach adaptors and generate amplifiable template after bisulfite conversion. Template is further extended with p5 and p7 adaptor and amplified.

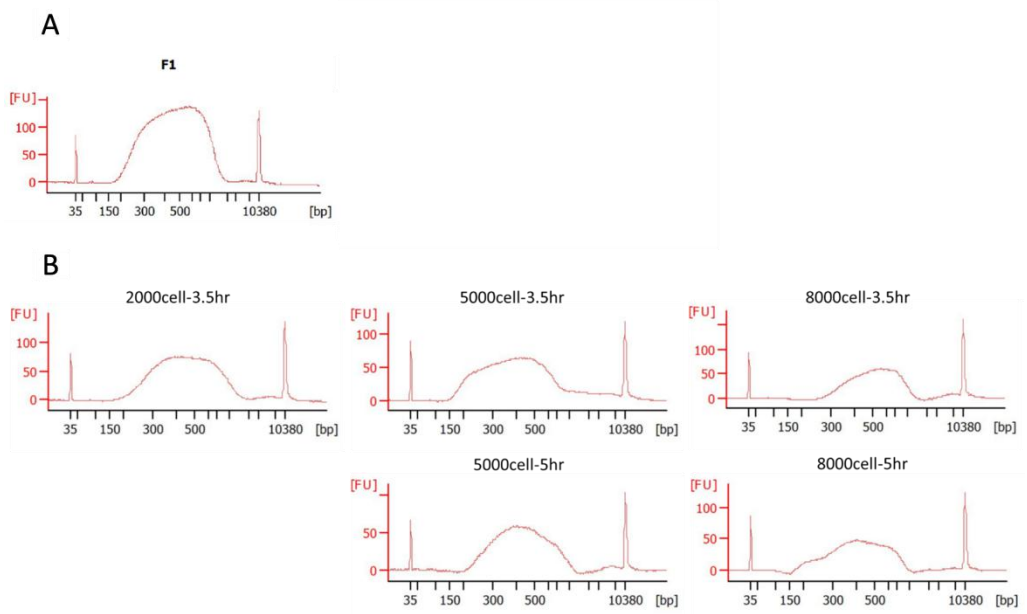


Figure 3-4 Representative size distribution of PBAT libraries. (A) is an exemplary size distribution of diluted libraries (Female pPGC sample 1), with majority of fragments within 200-1000bp (94%) and is similar to size distribution shown by Clark et al. (B) Different bisulfite incubation time and amount of starting material does not largely improve the size distribution.

3.3.3 General description of gonadal pPGC methylome

3.3.3.1 Thorough DNA demethylation occurs at all genomic features and regions in Wk5 pPGCs

Consistent with the LC-MS/MS results (Figure 2-33 (C)), the methylation level in WK5 pPGCs is extremely low (~1%) compared to somatic cells (median level over 75%) (Figure 3-5 (A)). Averaged CpG methylation level across all genes and neighbouring regions (± 5kb) in pig gonadal somatic cells are similar to human somatic cells (Tang et al., 2015, Gkountela et al., 2015) and also adult pig somatic lineages (Schachtschneider et al., 2015)(Figure 3-5 (B)). However, the methylation level in pPGCs is lower compared to human Wk7 PGCs (median 4.5%) (Tang et al., 2015) and E13.5 mPGCs (3.5% mCG) (Kobayashi et al., 2013). The reason for such low methylation level in pPGCs

remains unclear. As our PBAT libraries are generated with the same protocol used for generating human PBAT methylome, and with more starting DNA, it is unlikely that the low methylation level in pPGCs is solely due to inappropriate conversion of 5mC after saturation of cytosine conversion. Pig genome are known to have relatively lower content of repetitive elements (~40%) compared to other mammalian species such as human and mouse (Groenen et al., 2012, Fang et al., 2012). Repetitive elements usually maintain higher DNA methylation level (Lister et al., 2009). Therefore, it is possible that the lower content of repetitive elements in pig genome contributes to the lower genome-wide DNA methylation level and less escapees (see section 3.3.43.3.4.1) were observed in Wk5 pPGC compared to hPGCs and mPGCs.

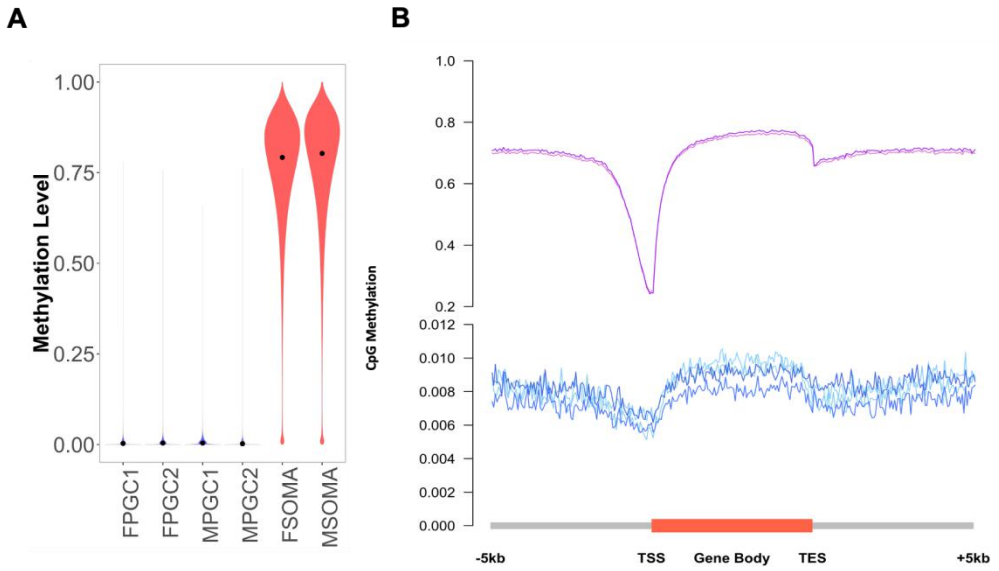


Figure 3-5 (A) CpG methylation levels in 1 kb genomic tiles of Wk5 (E35) female and male pPGCs and gonadal somatic cells. Black point indicates median. The median methylation level in pPGCs is around ~1%. (B) Averaged CpG methylation level profiles of all genes from 5 kb upstream (-) of transcription start sites (TSSs), through scaled gene bodies to 5 kb downstream (+) of transcription end sites (TES). Different y axes are used for pPGCs and somatic cells due to extreme low level of methylation in pPGCs.

Similar to previous reports in hPGCs (Tang et al., 2015), this thorough DNA demethylation in pPGCs occurs over all genomic features, including promoters and gene body regions which are highly methylated in somatic cells (Figure

3-6). As expected, most differentially methylated regions (DMRs) of imprinted genes (except PEG10), promoter regions of X-linked genes, and previously reported methylation-sensitive genes (Tang et al., 2015, Hill et al., 2018) are fully demethylation (Figure 3-7 and Figure 3-8), suggesting the DNA demethylation in PGCs is likely to be completed at this stage. Also, no clear difference is observed between male and female pPGCs. Note that in pig gonadal somatic cells, low- and intermediate-CpG density promoters are mostly highly methylated. This is different compared to human gonadal somatic cells where a large portion of those promoters show low or moderate methylation level (Tang et al., 2015) and reflects a potentially different regulation of gene expression. The reason for this difference is unclear. As I focus on PGCs this is not further investigated.

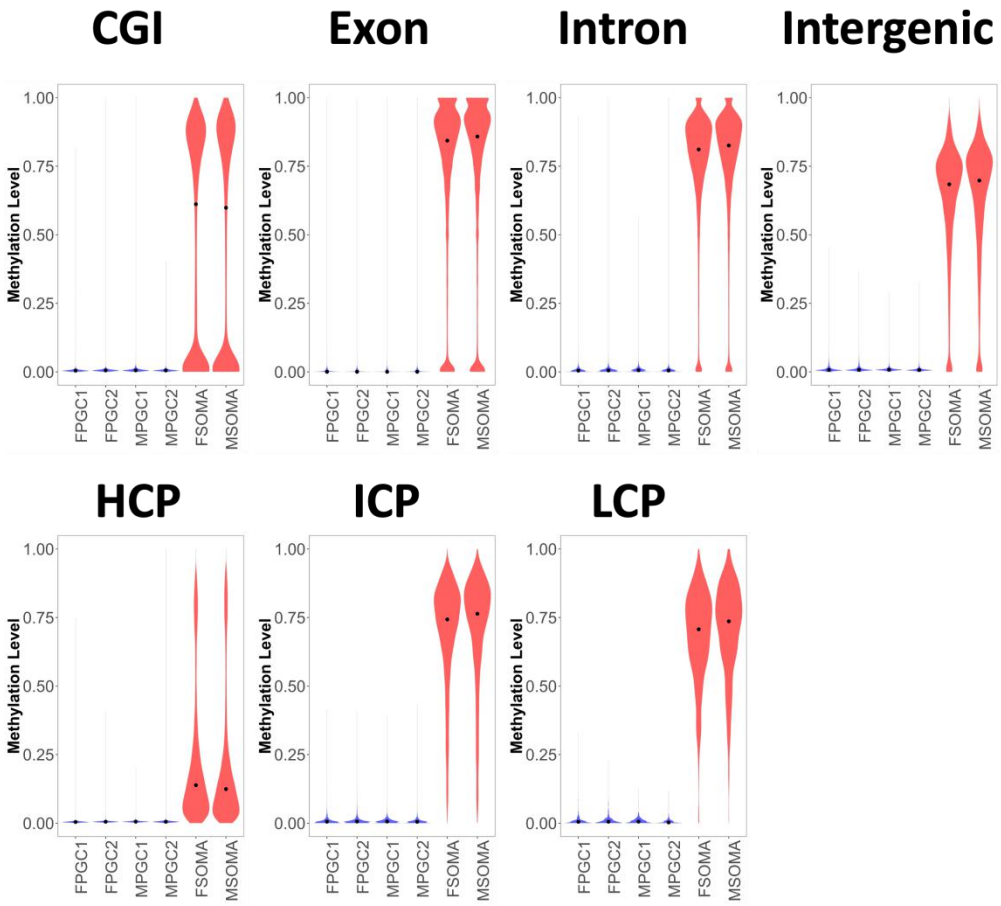


Figure 3-6 Violin plots showing CpG methylation levels in different genomic features in male and female Wk5 pPGCs and gonadal soma. CGI: CpG island;

HCP, ICP and LCP: high-, intermediate-, and low-CpG density promoters. DNA methylation is at basal level in all genomic features in pPGCs.

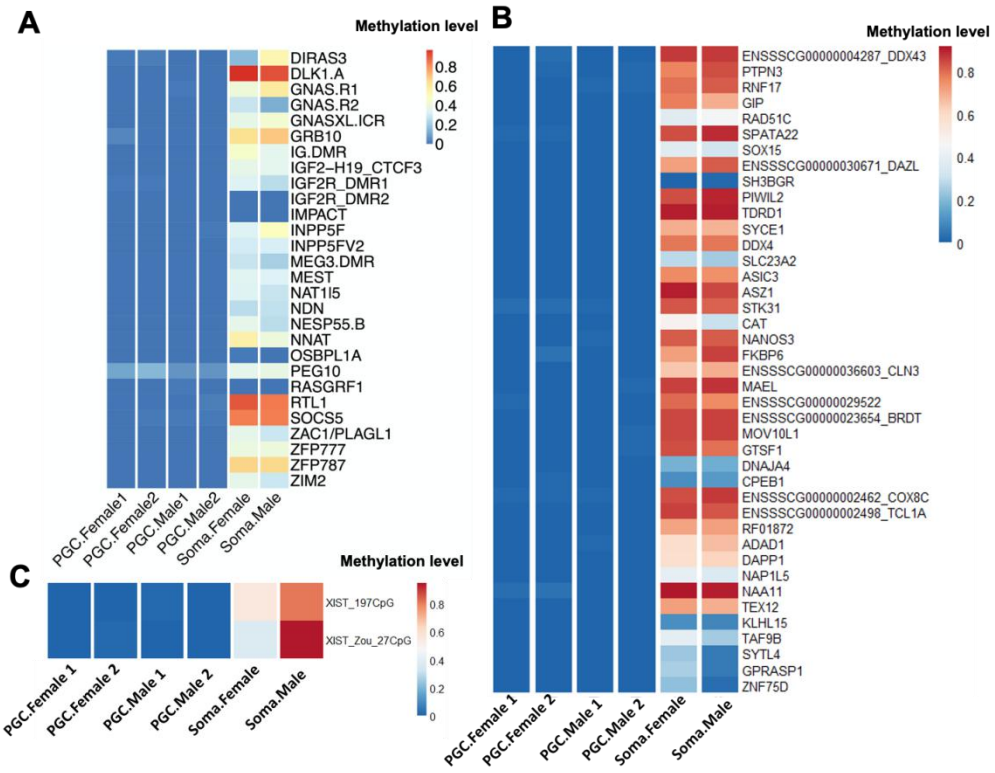


Figure 3-7 CpG methylation levels of (A) imprinted DMR regions, (B) promoters of previously reported methylation sensitive genes in germline as well as (C) whole length (197 CpG) and DMR (27 CpG) of XIST gene (Zou et al., 2019) in pPGCs and somatic samples. All examined regions, apart from the DMR of PEG10, are extremely demethylated in pPGC samples.

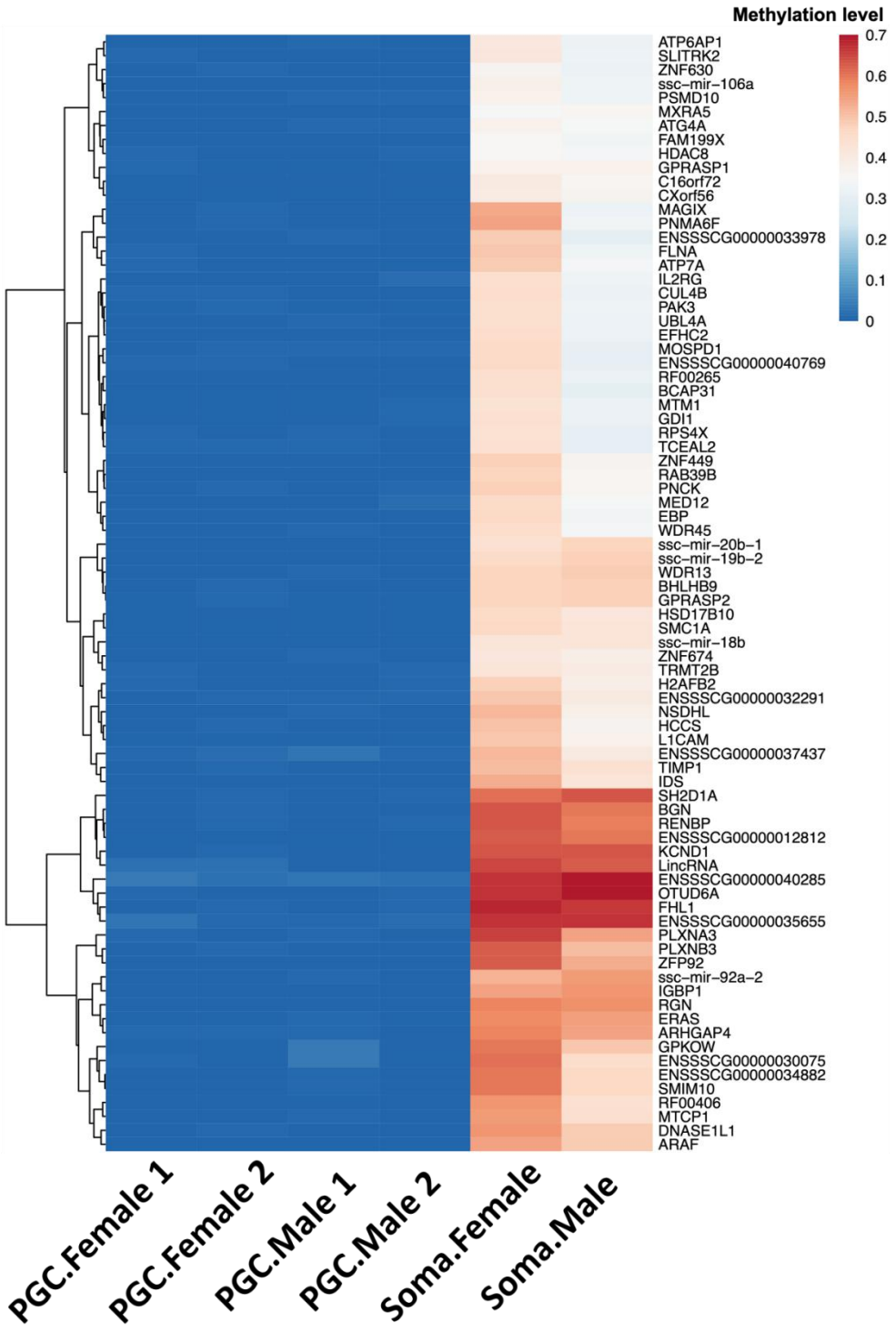


Figure 3-8 CpG methylation levels of CGI-containing X chromosome promoters. Promoters that are partially methylated (30%–70% methylation) in gonadal somatic cells are selected. Those examined regions are also demethylated in pPGC samples.

With the help of bioinformaticians, PBAT data of gonadal hPGCs (Wk5.5, Wk7 and Wk9) and mPGCs (E10.5 and E13.5) were processed and compared to Wk5 pPGCs. As expected, hPGCs show basal methylation level at Wk7-9 while mPGCs at E13.5. When global DNA methylation reaches basal level, PGC of all three species still show some retain of methylation, especially at TEs, suggesting the existence of demethylation-resistant loci (Figure 3-9). As the DNA methylation level of Wk5 pPGC and E13.5 mPGC is lower than hPGC of equivalent stages, it is not surprising that less demethylation-resistant tiles, with lower methylation level, were observed in pig (19,797) and mouse PGC (50,804) compared to hPGC (257,658)(Figure 3-9 (A)).

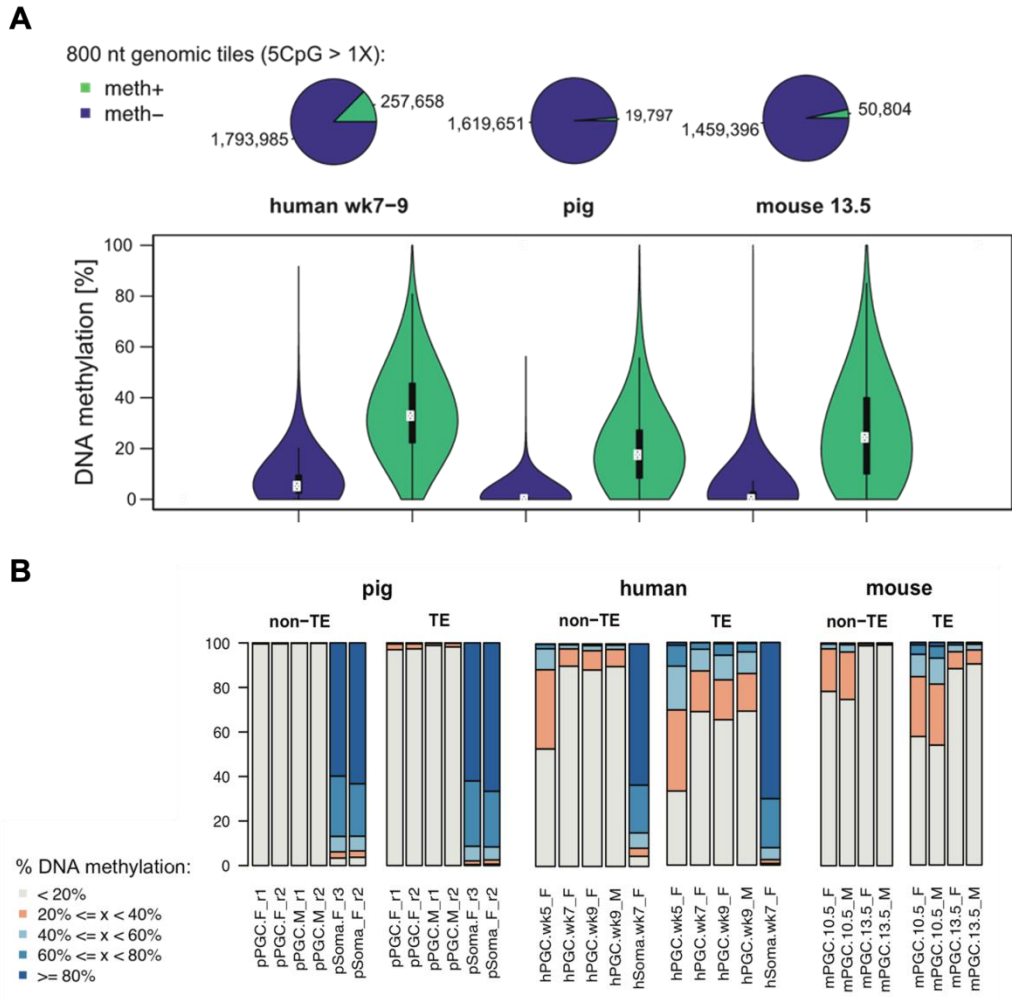


Figure 3-9 (A) Top: Proportion of demethylated loci (meth-) and demethylation-resistant loci (meth+) in Wk5 pPGCs, Wk7-9 hPGCs and E13.5 mPGCs (the number of meth+ and meth- 800nt genomic tiles are indicated on

the pie chart). Bottom: CpG methylation levels of meth- and escapees (meth+) in three species. White dot indicates median, black bar indicates interquartile range. (B) Distribution of CpG methylation in non-TE and TE genomic tiles (800nt, with at least 5 CpG with 1x coverage). TE regions show more residual methylation level compared to non-TE regions in all three species.

3.3.3.2 Methylation profile and upregulated expression of TEs in pPGCs

TEs, which make up almost all repetitive sequences in mammalian genome, can replicate, mobilize and insert themselves in a genome (Platt et al., 2018). Therefore, the activities of TEs have large impact on genome stability and evolution. DNA methylation is an important part of the host genome's defence mechanism against TE activities (de Mendoza et al., 2018, Jones and Takai, 2001). Here I focus on major families of retrotransposon as they make up the majority of TEs.

As in the previous section DNA methylation in gonadal PGCs is shown to be preferably retained at TEs, the methylation status of TEs in pPGCs is further investigated. All major families of TE show extreme low level of methylation in Wk5 pPGC (Figure 3-10 (A)) and surprisingly the expression of those TE families also start increasing in pPGCs from E14 compared to E14 somatic cells and E11 Epiblast (Figure 3-10 (B)), suggesting derepressed TE expression is concurrent with the global DNA demethylation in pPGCs. Still, TE activities are likely to be under control as other post-transcriptional repression mechanisms, including piRNA pathway, are activated in pPGCs (Table 3-1). In hPGCs, TE expression was found higher from Wk4 compared to ICM in one study (Guo et al., 2015). However, most TEs, apart from some evolutionary younger ones such as SVA and AluY, show lower expression in gonadal hPGCs compared to gonadal somatic cells (Tang et al., 2015), which appear to be different to pPGCs.

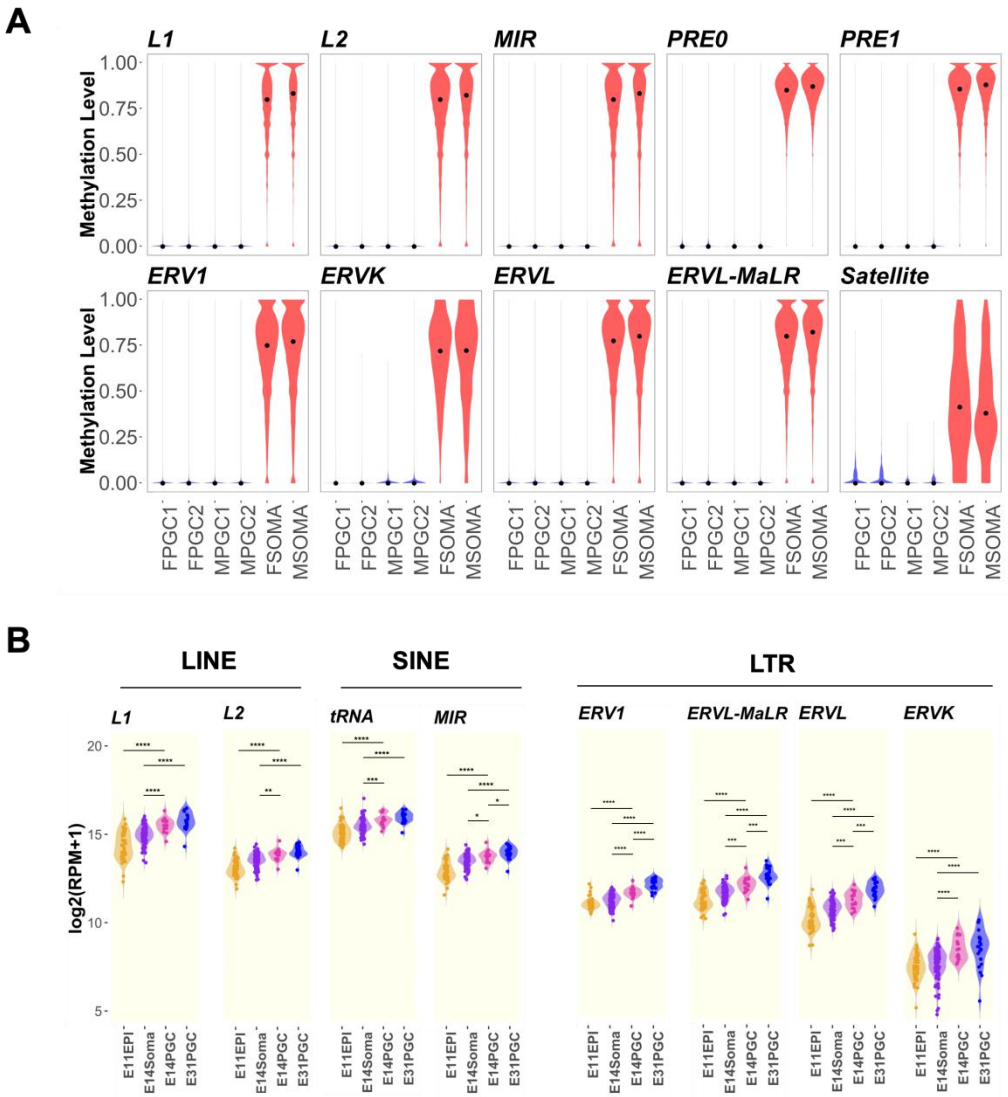


*Figure 3-10 (A) Violin plots showing CpG methylation levels at different TE families. PGC samples show basal level of methylation in all examined TE families (B) Expression profiles of major retrotransposon families in E11 epiblast, E14 somatic cells, E14 and E31 PGCs. p value. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$ by pairwise Wilcoxon test. Upregulation of retrotransposons is observed from E14 PGC.*

Table 3-1 Some of the TE repressor show increased expression (i.e. p_val_adj <0.05) in E14 and E31 pPGCs (pct.1) compared to E14 Soma (pct.2)

	p_val	avg_logFC	pct.1 (PGC)	pct.2 (E14 Soma)	p_val_adj	Gene Name	Note	Reference
ENSSSCG00000006846	3.63E-20	3.877893178	0.949	0.191	9.39E-16	HENMT1	piRNA/RNAi-Pathways	(Lim et al., 2015)
ENSSSCG00000026333	6.41E-16	3.787403539	0.795	0.124	1.66E-11	GTSF1	piRNA/RNAi-Pathways	(Yoshimura et al., 2009).
ENSSSCG00000006776	1.93E-14	1.692433662	1	0.854	4.99E-10	MOV10	Post-transcriptional	(Arjan-Odedra et al., 2012)
ENSSSCG00000016627	3.34E-13	5.144671815	0.564	0.022	8.65E-09	ASZ1	piRNA/RNAi-Pathways	(Ma et al., 2009)
ENSSSCG00000004752	9.68E-08	2.181766082	0.487	0.079	0.002504418	EXD1	piRNA/RNAi-Pathways	(Yang et al., 2016)
ENSSSCG00000016913	1.22E-06	3.77342896	0.282	0.011	0.031548459	DDX4	piRNA/RNAi-Pathways	(Kuramochi-Miyagawa et al., 2010)
ENSSSCG00000017543	1.33E-06	1.047803183	0.923	0.663	0.034394445	CALCOCO2	Post-transcriptional	(Guo et al., 2014b)
ENSSSCG00000030228	1.65E-06	2.914244046	0.359	0.045	0.042732316	PIWIL2	piRNA/RNAi-Pathways	(De Fazio et al., 2011)
ENSSSCG00000021075	4.66E-06	3.117193218	0.256	0.011	0.120706176	TDRD1	piRNA/RNAi-Pathways	(Reuter et al., 2009)
ENSSSCG00000039878	4.39E-05	3.202719741	0.179	0	1	PLD6	piRNA/RNAi-Pathways	(Watanabe et al., 2011)
ENSSSCG00000027770	4.66E-05	2.992331639	0.436	0.146	1	TDRKH	piRNA/RNAi-Pathways	(Saxe et al., 2013)
ENSSSCG00000027806	0.00015668	1.423693788	0.641	0.36	1	SAMHD1	Post-transcriptional	(Hu et al., 2015)
ENSSSCG00000007702	0.00016529	2.671545879	0.154	0	1	FKBP6	piRNA/RNAi-Pathways	(Xiol et al., 2012)
ENSSSCG00000011937	0.0002727	3.282425126	0.179	0.011	1	MORC1	Epigenetic	(Pastor et al., 2014)
ENSSSCG00000002552	0.00038579	1.531839852	0.256	0.045	1	TDRD9	piRNA/RNAi-Pathways	(Shoji et al., 2009)
ENSSSCG00000006645	0.00038583	0.770534126	0.923	0.921	1	SETDB1	Epigenetic	(Karimi et al., 2011)

ENSSSCG00000009570	0.00052254	0.470926685	0.949	0.933	1	Mphosph8	Epigenetic	(Robbez-Masson et al., 2018)
ENSSSCG00000032453	0.0007548	1.333759652	0.513	0.281	1	SUV39H1	Epigenetic	(Bulut-Karslioglu et al., 2014)
ENSSSCG00000009749	0.01360181	2.587395704	0.103	0.011	1	PIWIL1	piRNA/RNAi-Pathways	(Reuter et al., 2011)
ENSSSCG00000014016	0.0494541	0.311946008	0.923	0.989	1	SQSTM1	Post-transcriptional	(Guo et al., 2014b)

Studies in hPGCs (Tang et al., 2015) and early embryo (Guo et al., 2014c) suggest that evolutionarily younger TEs retain higher level of methylation residual during epigenetic reprogramming. As the overall methylation level of major TE families are extremely low and no clear difference can be observed, the composition of TEs that retain partial methylation (i.e. methylation level >=10%) was further examined. Long interspersed elements 1 (LINE1 or L1), the younger family of LINE, is shown overrepresented in pPGCs compared to somatic cells (Figure 3-11 (A)). Similarly, relatively larger proportions of younger, pig specific TEs, such as L1_SS and Pre_0SS, show more retain of DNA methylation (i.e. methylation level >=15%) (Figure 3-11 (B)). These suggest, despite the overall low methylation level in TEs, DNA methylation is preferably retained at evolutionary younger and potentially active TEs (Groenen et al., 2012)) in pPGCs.

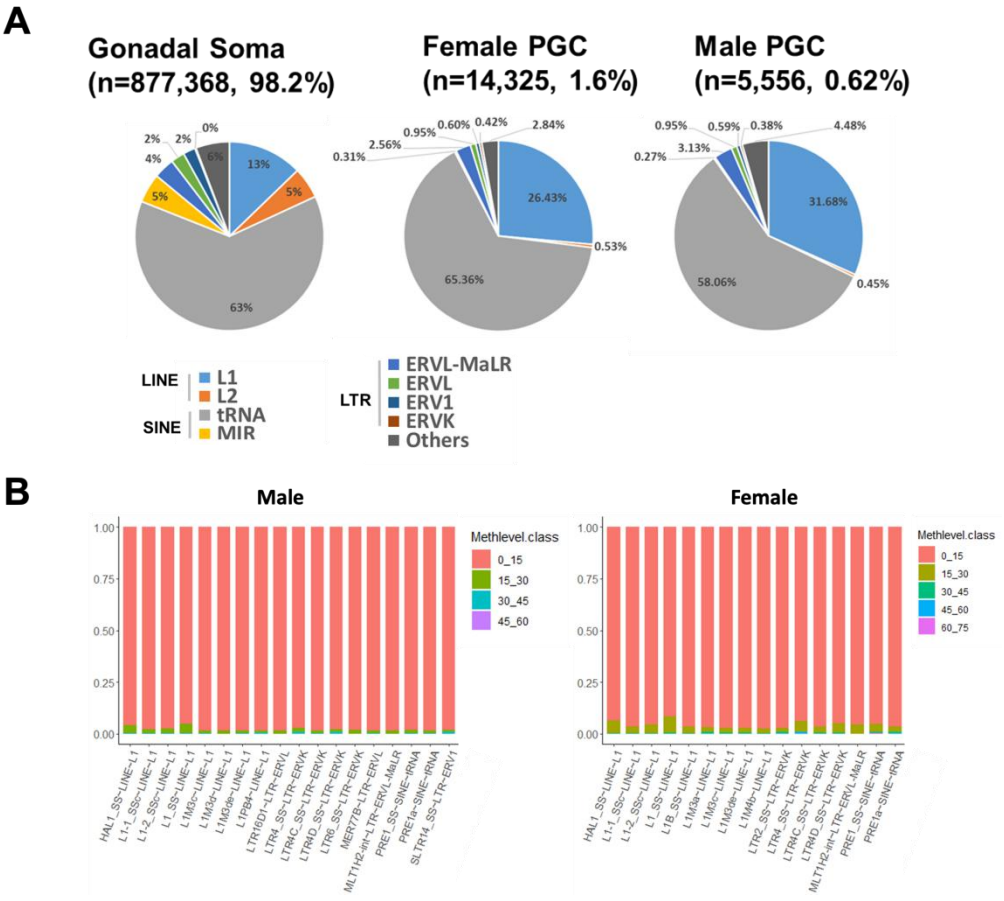


Figure 3-11 (A) Distribution of major TE families that retain partial methylation (≥10%) in pig gonadal soma and PGCs. N: indicates the number

of TEs that retain partial methylation, followed by the percentage of those partially methylated TEs among all TEs of the families indicated. L1 is overrepresented in pPGCs samples. (B) Distribution of CpG methylation in the top demethylation-resistant retrotransposons (i.e. retrotransposons with the lowest percentage below 15% methylation) in male and female pPGCs. X axis shows the Name-Family-Subfamily of the retrotransposons. Evolutionary younger and pig-specific retrotransposons such L1_SS, PRE1_SS and LTR4_SS show higher percentage of copies that retain partial methylation.

3.3.4 **Escapee in pPGCs show similar and distinct features compared to other species**

3.3.4.1 Identification of escapees

After confirming the extreme low level of DNA methylation in Wk5 pPGCs, Wk7-9 hPGCs and E13.5 mPGCs, significantly hypermethylated regions (HyperMRs) in PGC of these three species were identified by *hmr* function of Methpipe (Song et al., 2013). These HyperMRs were further filtered by size (>= 5 CpG) methylation level to identify high-confidence ones, which were referred as “escapees” in the following analysis. As the basal levels of DNA methylation in PGC are different among three species, different methylation thresholds (at least 30% for human and 15% for pig and mouse) were selected based on distribution of residual methylation level (Figure 3-9 (A)) to ensure that the escapees in different species are comparable. (i.e. If 30% methylation was used for filtering pig HyperMRs, less than 23% of HyperMRs will be identified as escapees (Figure 3-12), while in hPGCs using 30% as threshold will return around 50% of HyperMRs as escapees.) As expected, less escapees are identified in pPGC (10,421) and mPGC (21,471) compared to hPGC (100,208) (Figure 3-13).

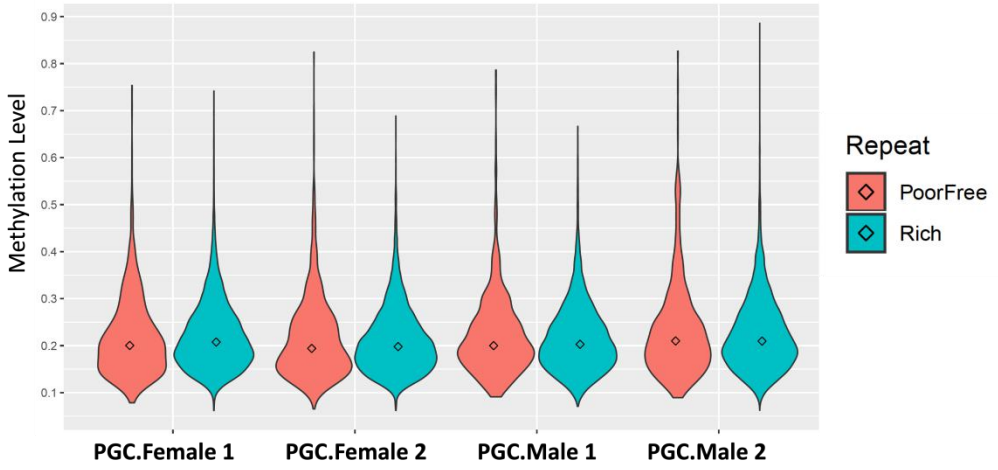


Figure 3-12 Violin plot showing the methylation level of HyperMRs in pPGC samples. TE-poor/free and TE-rich HyperMRs.do not show major difference. Diamond indicates median methylation level, which is around 20% in all PGC samples.

3.3.4.2 Characterisation of escapees in human, pig and mouse PGCs

The features of the escapees in three species were further investigated. In all three species, majority of escapees (over 90% in all three species) are found with at least 10% overlapping with TEs (i.e. TE-rich escapees) (Figure 3-13). Compared to TE-poor escapees, TE-rich escapees show enriched overlaps with intronic, intergenic or distal regions, which is possibly due to a biased distribution of TEs (Figure 3-13). It is estimated that, among the TE insertions into human genes, the majority (~89.5%) locate within introns (Nekrutenko and Li, 2001). The proportion of TE insertions into the exons are also very low in pig (<1%) (Chen et al., 2019a) and mouse genome (Nellåker et al., 2012, Sharif et al., 2013). Also, it has been suggested that the distribution of younger Alu and LINE1 in human genome is biased toward AT-rich regions which are by definition GC-poor and usually has low gene density (Costantini et al., 2011, Sultana et al., 2019). More TE-rich escapees in pig and human overlap with intronic regions compared to in mouse. The reason for this difference is unknown but it is possible that this enrichment of intronic TE-rich escapees in

pig and human may reflect a different distribution of TE or TE repression mechanisms compared to in mouse. Notably, the most abundant repeat families at TE-rich (>=10% overlap with TE) escapees are species-specific and evolutionarily young TE, including the pig SINE element Pre0_SS of the PRE1-family, human AluY, and mouse IAPEz repeats (Figure 3-14), which is consistent with the description in section 3.3.33.3.3.2 and suggests a conserved mechanism in mammalian species that maintains partial methylation at those potentially active TEs.

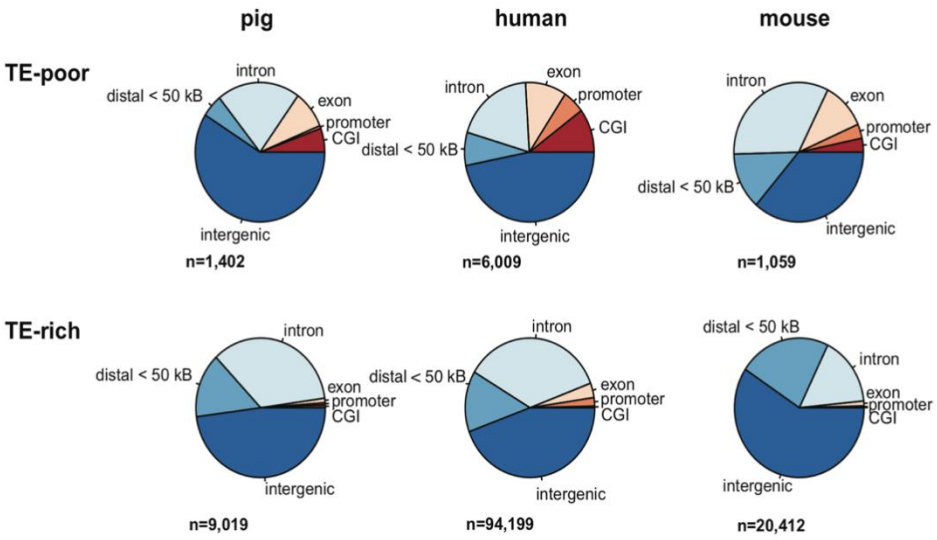


Figure 3-13 Distribution of TE-poor (<10% overlap with TE) and TE-rich (≥10% overlap) escapees in Wk5 pPGCs, Wk7-9 hPGCs and E13.5 mPGCs. Escapees (n) are high confidence HyperMRs that were filtered by methylation level (at least 30% in human and 15% in pig and mouse).

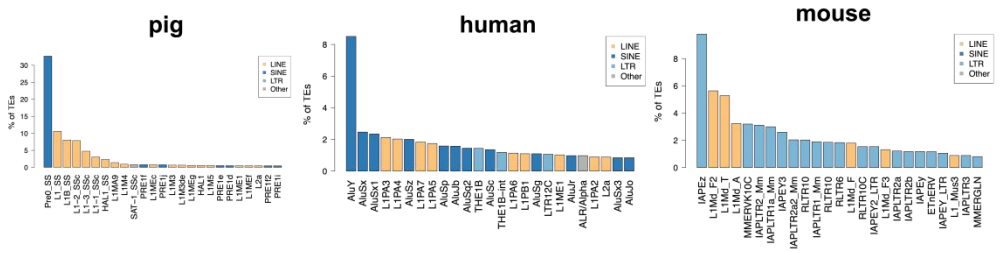


Figure 3-14 Distribution of TE families that overlap with TE-rich escapees in Wk5 pPGCs, Wk7-9 hPGCs and E13.5 mPGCs. Evolutionary younger TEs, such as Pre0_SS and L1_SS in pig, AluY and L1PA in human as well as IAPEz in mouse, represent higher percentage

A larger portion of TE-poor (<10% overlap with TE) escapees (10%) is identified in pPGC, compared to less than 6% in hPGCs and less than 5% in mPGCs (Figure 3-13). This is potentially also because the relative content of repetitive elements is lower in pig genome (Groenen et al., 2012). Compared to TE-rich escapees, TE-poor escapees in all three species show more abundance at CGIs, promoters and exons. Still, differences such as underrepresentation of promoter in pig and higher proportion of intron in mouse, were observed between species (Figure 3-13). TE-poor escapees were than compared between three species at sequence level. 208 out of 1059 TE-poor regions in the mouse, and 205 out of 1402 TE-poor regions in the pig can be lifted over to human (i.e. can be identified as syntenic regions to human genome). There is little conservation at sequence level between mouse and the other two species (Figure 3-15). More syntenic, TE-poor escapees in pig show intersection with human TE-poor escapees, which compose a significant 21.5% of all TE-poor escapees in pig (Figure 3-15). When comparing TE-poor escapees at homologous gene level, more conservation was observed. 265 (47%) escapee genes in pig and 191 (23.2%) in mouse are in common with human escapee genes (Figure 3-15).

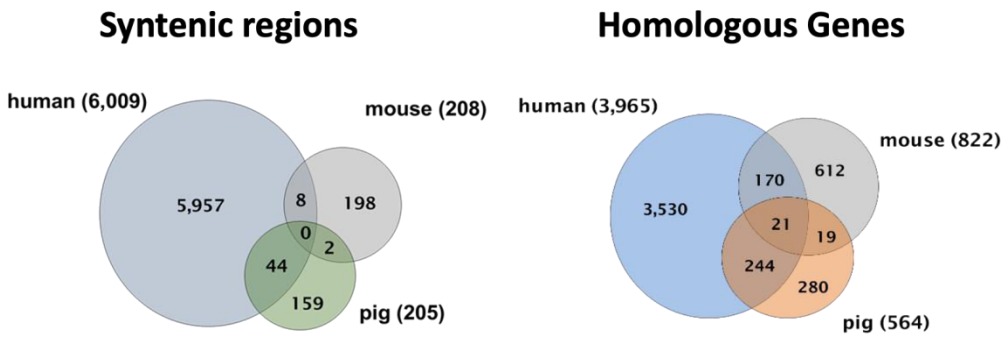


Figure 3-15 (Left) Overlap of syntenic TE-poor escapees among pig, human and mouse. Escapee regions in pig (205) and mouse (208) were lifted over to human genome and compare with human escapee regions. (Right) Overlap of homologous, TE-poor escapee genes among pig, human and mouse.

Consistent with previous report in hPGC (Tang et al., 2015), comparison with the NHGRI GWAS catalogues reveals that the 265 human-pig conserved escapee genes are linked with metabolic and neurological traits and diseases such as "obesity-related traits", "schizophrenia" and "breast cancer" (Figure 3-16). Some of those disease-associated escapees also show conservation between human and pig at sequence level, such as obesity-related gene *SORCS2* and schizophrenia-related *PLCH2 (*Figure 3-17*)*. Similar terms are linked with the 191 human-mouse conserved escapee genes, albeit with different genes associated to those terms (Figure 3-16). For the species-specific escapee genes, the comparison with GWAS catalogue reveal common as well as species-specific associations, such as common association with "Adolescent idiopathic scoliosis", and pig-specific association with "Asthma" why may be worth further exploration (Krauss-Etschmann et al., 2015) (Figure 3-18).

Functional enrichment analysis was also performed for escapee genes which are detected in human and pig. These common escapee genes show enrichment for KRAB, ZnF_C2H2 and SCAN protein domain (Figure 3-19), which are key domains in members of KRAB-ZFP family (Ecco et al., 2017). The common escapees are also enriched for brain-specific expression, consistent

with their association with neurological-related traits (Figure 3-19).

Interestingly, demethylation-resistant, pig-specific TEs also cover genes enriched for developmental-, metabolic- and neurological-related GO term, such as *FTO*, which is an obesity-related gene and covered with Pre_0SS (Figure 3-17 and Figure 3-20). This suggests that retained methylation on single-copied genes and species-specific TEs may contribute to evolutionarily conserved and also species-specific epigenetic inheritance.

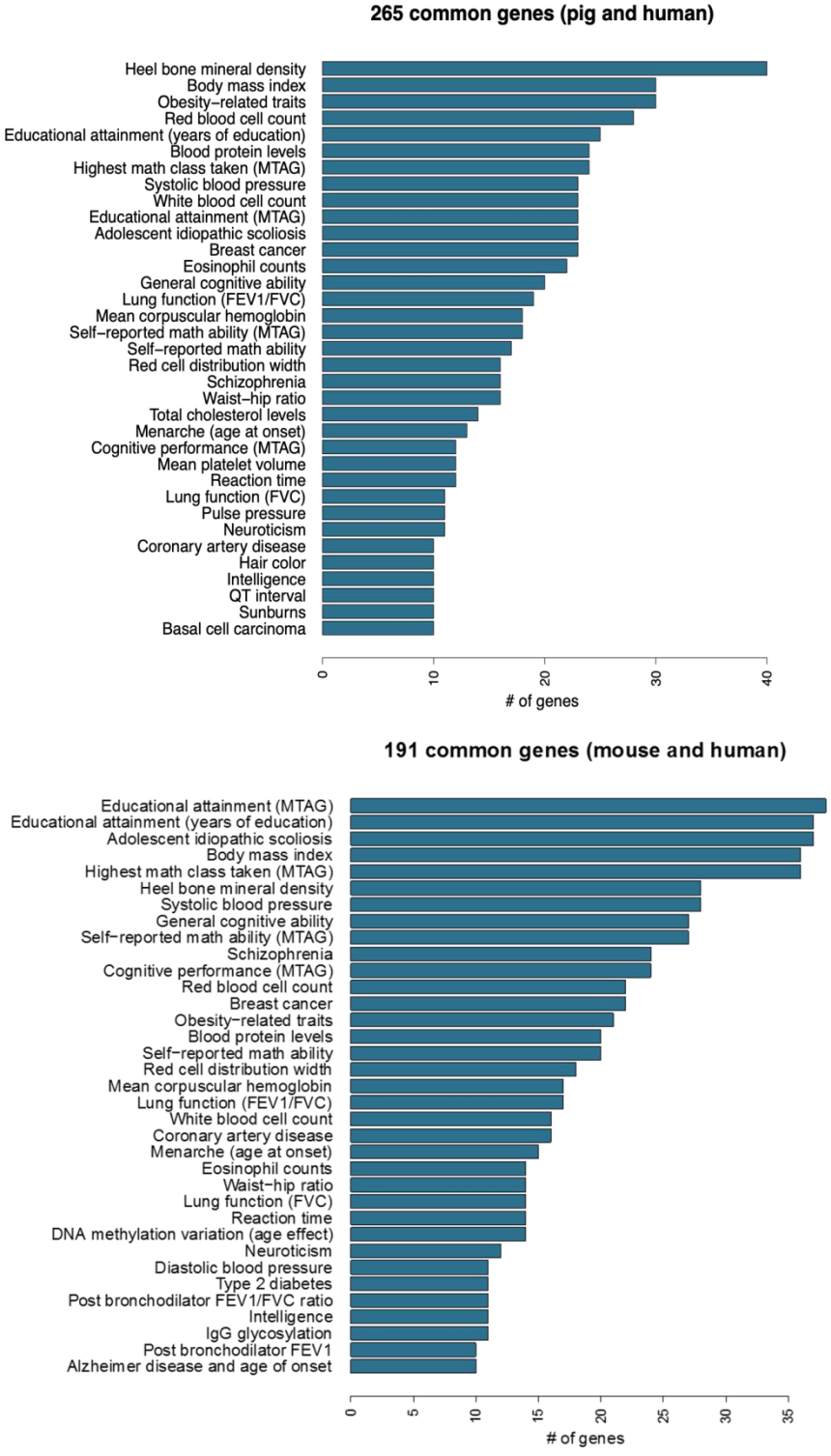


Figure 3-16 Enrichment analysis for GWAS catalogue terms showing disease and traits associated with conserved TE-poor escapee genes between pig and human (top) and mouse and human (bottom). Several common terms such as

"breast cancer" and "Schizophrenia" are identified in both groups, albeit with different escapee genes associated.

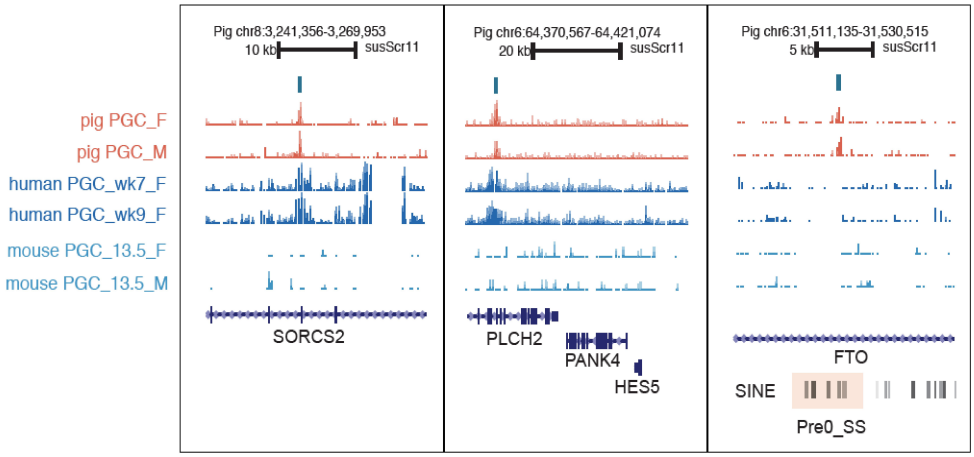


Figure 3-17 The TE-poor escapee regions within SORCS2 and PLCH2 are conserved between human and pig, while a pig-specific escapees, overlapping with SINE element Pre_0SS, is identified within FTO.

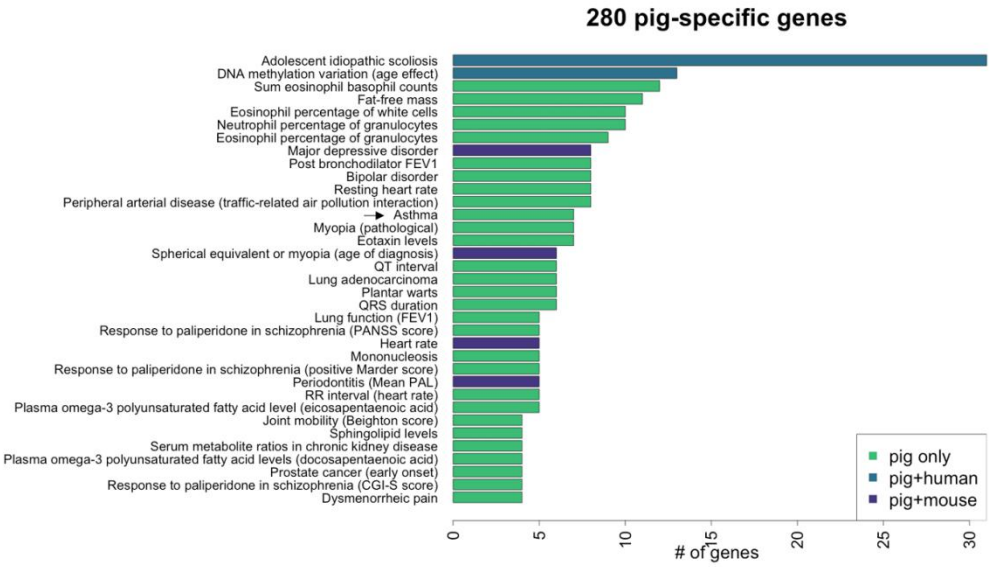


Figure 3-18 Enrichment analysis of GWAS catalogue terms for pig-specific TE-poor escapee genes. Black arrow point to the term "Asthma" that was mentioned in the text. Pig-only and shared terms with human or mouse were labelled with different colours. Note that here only the top significant 35 terms

were compared across species, and therefore some of the "pig only" terms were also significant in other species but not within the top 35.

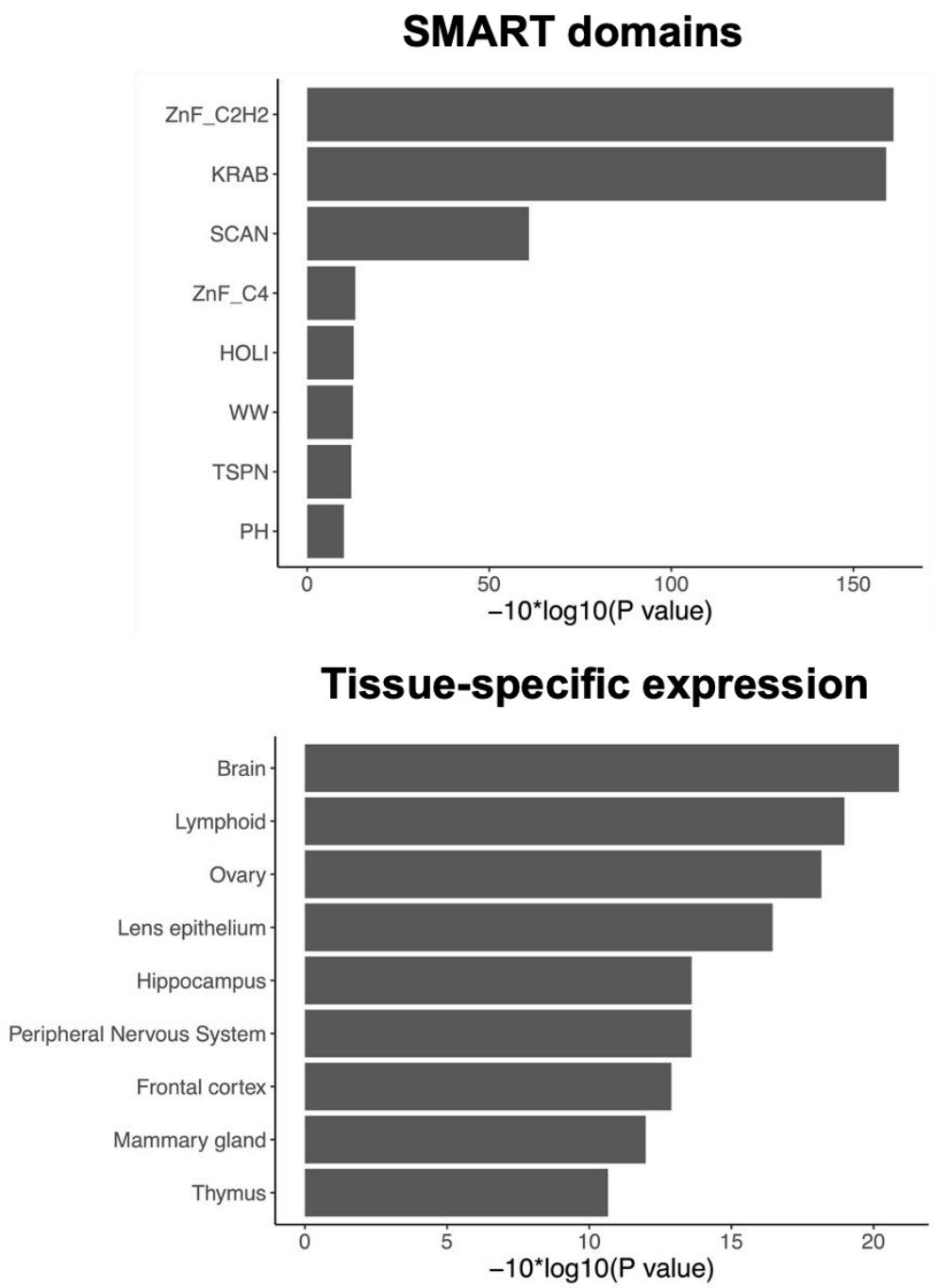


Figure 3-19 Enrichment analysis for SMART domains and tissue-specific expression (UniProt UP_TISSUE) of TE-poor escapee genes shared by human and pig.

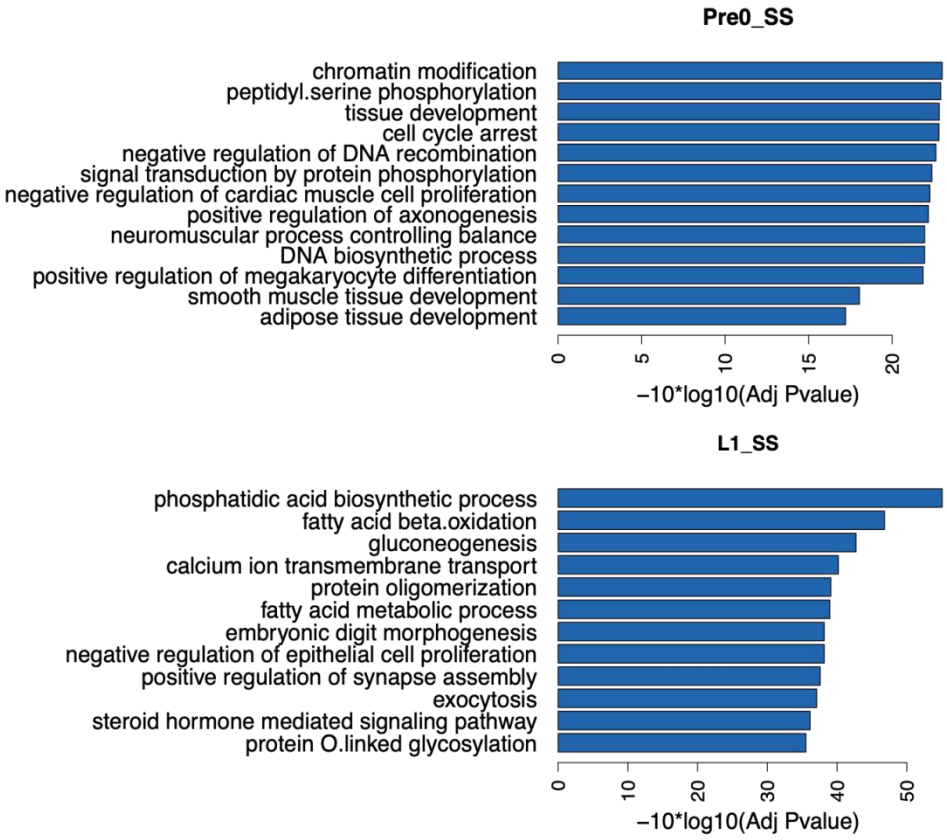


Figure 3-20 GO term enrichment analysis of genes overlapping demethylation-resistant pig-specific TE, Pre_0SS (top) and L1_SS (bottom). Interesting terms such as "positive regulation of axonogenesis" and "fatty acid beta oxidation" were identified.

3.3.5 **Methylation profiles of the TE-poor escapee regions in pig gonadal somatic cells and blastocysts.**

Escapee regions have the potential to be the carrier of epigenetic information for inter- and transgenerational inheritance. However, if and how escapee regions are selected and maintained during epigenetic reprogramming in PGCs, gametes as well as early embryos remain unknown.

By definition escapee regions have higher methylation level compared to neighbouring regions. However, it is not impossible that hypermethylation of those regions are a common feature among different cell types (i.e. During global DNA demethylation, if methylation level is reduced equally across

genome, previously hypermethylated regions can be perceived as escapees in gonadal PGCs.). Therefore, the methylation profiles of high-confidence, TE-poor escapee regions (i.e. escapee regions are common among all male and female pPGC samples, n=133) as well as their up- and downstream regions are investigated in pig gonadal somatic cell. Those regions show similar methylation status in male and female somatic cells, and in contrast to the profile in pPGCs, methylation level of a large portion of these regions do not show higher methylation level in pig gonadal soma compared to neighbouring regions (Figure 3-21). This suggests these TE-poor escapees are specific to pPGCs, and therefore might be selectively and actively maintained in PGC.

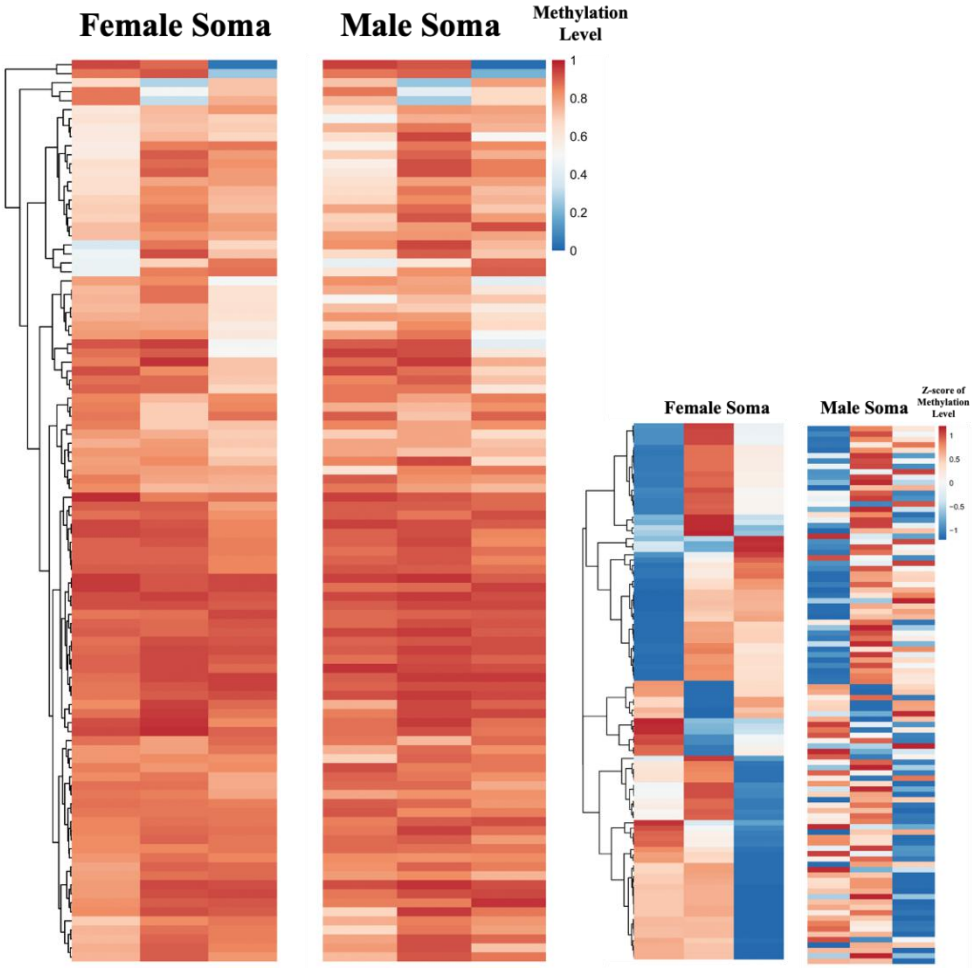


Figure 3-21 (Left) CpG methylation level at escapee and its neighbouring regions (i.e. upstream and downstream regions with the same length as the escapee) in pig gonadal somatic samples. (Right) Z-score of methylation level showing that in gonadal soma, a large portion of the escapee regions do not

show higher methylation level compared to neighbouring regions, suggesting the hypermethylation of escapees is the unique feature of PGCs.

Majority of escapees identified in gonadal hPGCs become high methylated in gametes and retain partial methylation in blastocyst. PBAT data in gametes are currently unavailable, but fate of TE-poor pig escapees can be investigated in early blastocyst (Canovas et al., 2017). The global methylation level of the available E7.5 pig in vivo blastocyst data is around 12%, suggesting the epigenetic reprogramming in those early blastocysts is finished while de novo methylation process has probably barely started (Zhu et al., 2018). Therefore, the E7.5 pig blastocyst data can be used for investigating the fate of escapees. As the processed PBAT data in blastocyst was aligned to previous pig genome assembly (susScr3), high-confidence TE-poor escapee regions (i.e. escapees that are common in at least two pPGC samples) that can be lift over to the susScr3 assembly (n=279) were selected. Similar to the previous report in human, majority of the TE-poor escapees also show partial methylation in blastocyst, suggesting that those escapees may have the ability to escape both epigenetic reprogramming events in the mammalian life cycle. Note that the methylation status in some selected escapee regions vary largely among the three replicates of blastocyst, which is also observed in hypomethylated human blastocyst and warrant further examination on a larger scale of how escapees are able to escape the global demethylation (Figure 3-22).

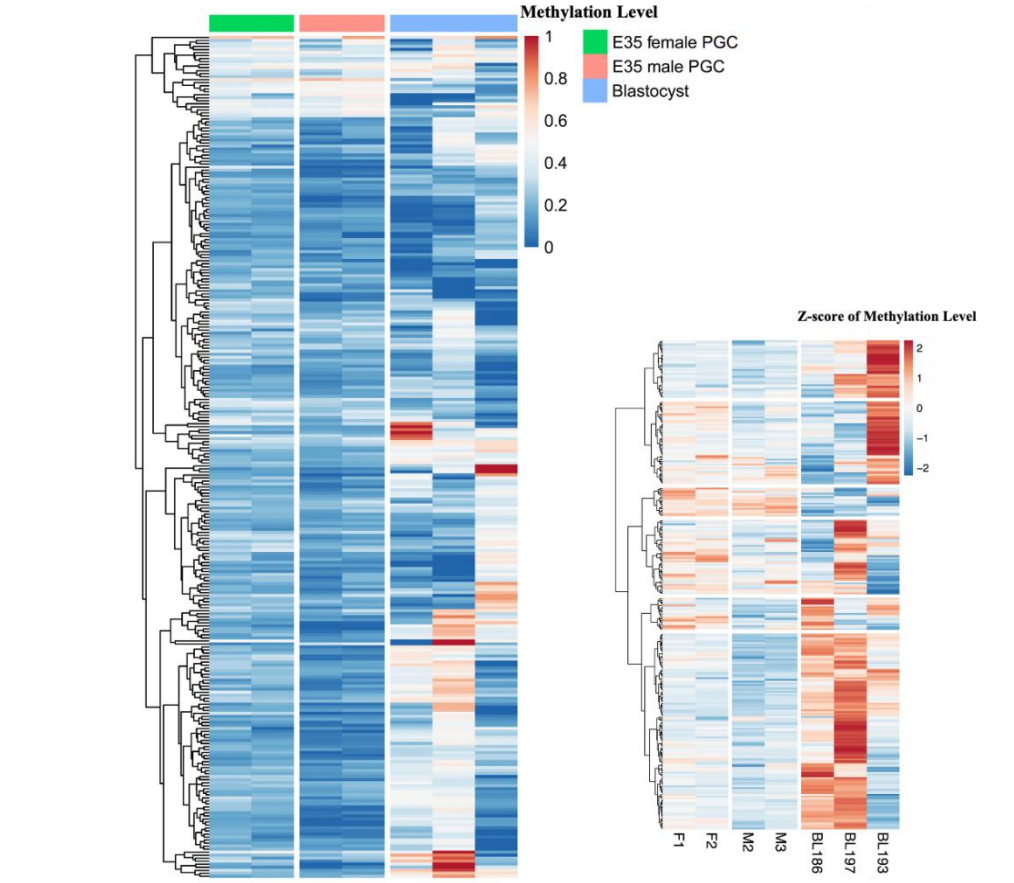


Figure 3-22 (Left) CpG methylation level at escapee in female and male pPGC and three blastocyst samples. Majority of escapees also show retain of methylation in blastocyst. (Right) Z-score of methylation level showing that for majority of escapees, compared to PGCs higher methylation level is observed in at least one blastocyst. However, the methylation level also varies largely among blastocyst sample.

3.4 Summary

In this chapter, base-resolution, high coverage DNA methylome is obtained using PBAT-seq. Analysis of this dataset reveals basal DNA methylation level in Wk5 pPGC, with methylation residuals observed at some repetitive elements and a small number of TE-poor, single-copy genes. These findings are consistent with previous reports in Wk7-9 hPGCs and E13.5 mPGCs, and therefore the methylome data of gonadal PGCs were further compared between pig, human and mouse.

Wk5 pPGC shows the lowest global DNA methylation level among three species. The reason for this remains to be investigated, but possibly this is due to (A) the lower content of repetitive elements in pig genome and (B) the high purity of pPGC population after double-FACS. Similar to gonadal hPGCs and mPGCs, methylation residuals exist in pPGCs when global DNA methylation reaches the basal level. Methylation residuals in PGCs of all three species are preferably maintained at repetitive elements, especially the evolutionary younger retrotransposons. Escapees which are regions show significantly higher methylation residuals compared to the neighbouring regions were then identified. Some of the escapee regions also show retention of DNA methylation in early pig blastocyst, which is similar to human escapees and suggests they have the potential to carry epigenetic information to next generations. As expected, in all three species over 90% of those escapees largely overlap with TEs (i.e. TE- rich), especially the species-specific TEs. A small portion of those escapees are TE-poor and cover single-copy genes, which showed conserved and divergent features across three species including conserved association to metabolic and neurological disorders and pig-specific association to respiratory disorders. Still, the TE-rich escapees also cover genes enriched for developmental-, metabolic- and neurological-related GO term and therefore may also contribute to species-specific epigenetic inheritance. Overall, it appears to be a conserved event in mammalian species that DNA methylation is retained at selected regions, such as evolutionarily young TEs and metabolic-related genes, during the global demethylation in PGCs. The underlying mechanisms and potential biological consequence, such

as transgenerational epigenetic inheritance, remain unclear and warrant further investigations.

Chapter 4 General Discussion and Further Work

In this project, with scRNA-seq for transcriptome and PBAT-seq for DNA methylome, unique features and events have been identified during the specification and development of pPGCs (Figure 4-1), which provides insights for PGC development in non-rodent species including human. Here, I further discuss the potential mechanisms for those events and their impact on germline specification and development.

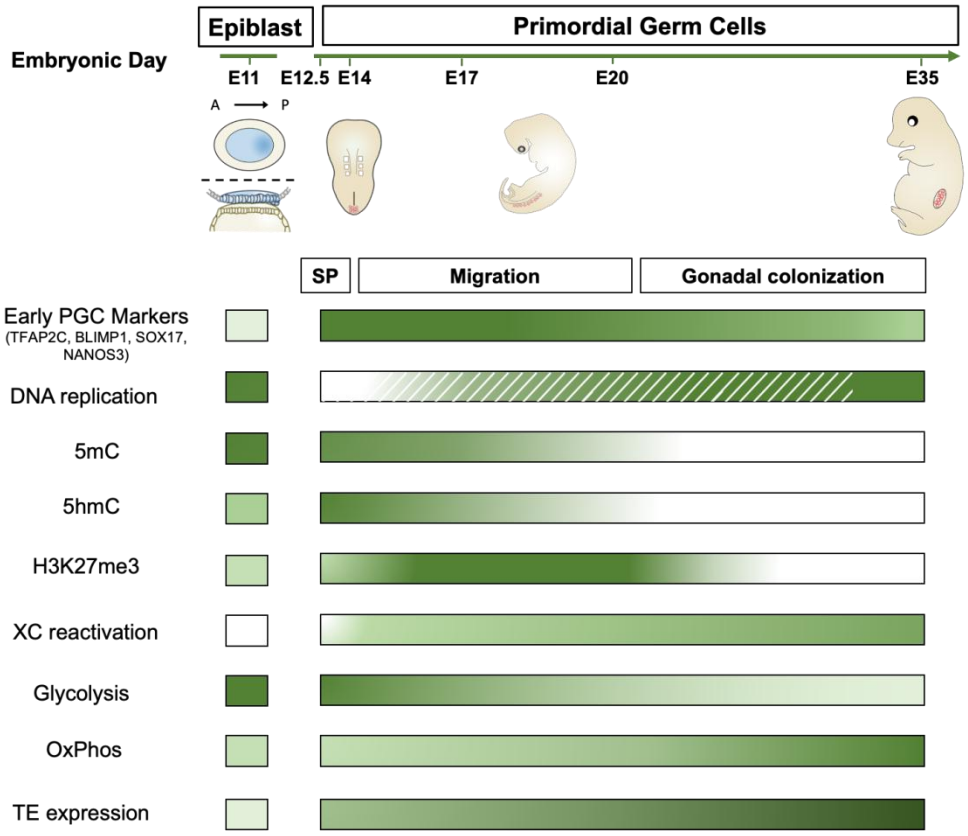


Figure 4-1 Diagrammatic representation of the events in the pig germline. SP: indicates germline specification; OxPhos: indicates oxidative phosphorylation; The dash line indicates expected DNA synthesis in migratory and early gonadal PGCs.

4.1 Dual origin of PGC in non-rodent mammals?

As described in Introduction (section Chapter 11.11.1.11.1.1.1 and Chapter 11.11.1.21.1.2.1), the specification of mPGCs is well-studied, of which only one origin is identified in the proximal epiblast of the pre-gastrulation, cylinder-shape embryos (Lawson and Hage, 1994, McLaren and Lawson, 2005). In vitro studies with human and monkey PSCs as well as in vivo study with pig embryos, which also develop into bilaminar disc before gastrulation, suggest that the posterior epiblast is likely to be the origin of PGCs in non-rodent mammals with disc-shape embryos (Kobayashi et al., 2017). However, recent studies of in vivo monkey embryos (Sasaki et al., 2016) and 3D human embryos cultured in microfluidics (Zheng et al., 2019) identified putative PGC precursors (i.e. $TFAP2C^{+}SOX17^{+}$ and $TFAP2C^{+}NANOG^{+}SOX17^{+}$ cells) located in the dorsal nascent amnion, suggesting another possible origin of primate PGCs. Regardless of the potential origin, determined primate PGCs which are BLIMP1-positive are found as a cluster around the gastrulating posterior epiblast (Sasaki et al., 2016, Zheng et al., 2019).

In this project, the pre-migratory pPGCs were isolated from the cluster at the posterior end of late gastrulating embryos, which were surrounded by presumably late primitive streak cells and amniotic precursors. Given this close spatial relationship, the possibility cannot be ruled out that PGC-competent cells in pig may also give rise to amniotic cells under BMP signalling (Yoshida et al., 2016, Valdez Magaña et al., 2014). However, unlike in primates, morphologically distinguishable amniochorionic tissue and amnion cavity are identified in pig after gastrulation when determined pPGC already exists (Hassoun et al., 2010). Therefore, studies of pPGC specification in vivo will provide insights in evolutionary perspectives, but to address the origin of hPGC in vivo non-human primate embryos and self-organised 3D human embryos in vitro will be required.

A recent review mentioned another possibility of dual origin of primate PGCs, suggesting that the nascent amniotic cells in primates may retain features of early epiblast cells that are required for germline competence and therefore under the right cues, i.e. WNT from the adjacent cytotrophoblast layer and BMP from amnion itself, those cells can give rise to PGCs (Kobayashi and

Surani, 2018). It is worth noting that, at transcriptional level, early primate PGCs are more similar to gastrulating cells (Ma et al., 2019, Tyser et al., 2020). Instead of inferring the origin of PGCs solely based on transcriptional similarity, single cell transcriptome combined with lineage tracing (Kester and van Oudenaarden, 2018) may be more useful for validating this possibility.
If the amniotic or dual origin of primate PGCs is proven true, this will lead to more questions. First of all, is the nascent amnion, instead of posterior epiblast, the primary, if not sole, origin of PGC? With in vivo embryos and in vitro microfluidic modelling, the germline competence will need to be characterised meticulously. Spatial transcriptomics and advanced bioinformatic tools such as RNA velocity analysis (Vickovic et al., 2019, van den Brink et al., 2020, La Manno et al., 2018) can be used to compare the nascent amnion, posterior epiblast and PGC precursors at molecular level and to understand their surrounding signals and transitions of cell identity. It is also possible that the PGC-competent state can be defined at epigenetic level, such as chromatin accessibility and bivalent promoters (Verma et al., 2018), which can be investigated with scNMT-seq (Clark et al., 2018) and CUT&Tag technique (Kaya-Okur et al., 2019). Second, how long is the germline competence retained in the nascent amnion and/or the posterior epiblast? Unlike previously suggested in new-specified pPGCs (Kobayashi et al., 2017), the nascent cyPGCs appear to be proliferative and the expansion of cyPGC cluster is achieved by both proliferation and recruitment from posterior amnion (Sasaki et al., 2016). This hints a potentially shorter time window for germline competence in primates than previously estimated based on pig model. Last, it is important to understand if and how those PGC precursors and/or nascent PGCs actively migrate down the amnion towards the posterior partial endoderm (PE), as ectopic PGCs may be removed via cell death (Stallock et al., 2003) or otherwise may cause detrimental consequences such as germ cell tumours (Schneider et al., 2001).

4.2 Unique epigenetic signatures may contribute to pPGC specification and development

Multiple epigenetic and transcriptional events that are unique to PGCs have been observed in this project. Here I mainly discuss the events observed in new-specified pPGCs, which may provide directions for further investigation about hPGC specification with the in vitro systems.

4.2.1 **5hmC enrichment in newly specified pPGCs**

The enrichment of 5hmC has been previously reported in newly specified pPGCs (Kobayashi et al., 2017) and in some loci of D4 hPGCLCs (Tang et al., 2015). In this project this enrichment has been further confirmed with both immunofluorescence and LC-MS/MS. E14 pre-migratory pPGCs exhibit higher 5hmC level compared to pluripotent cells (E11 Epiblast) and the surrounding somatic cells and the D4 hPGCLC shows a similar trend when comparing to hESCs and in vitro somatic lineages. As expected, 5hmC is reduced to almost absence in gonadal hPGCs and pPGCs. These results suggest that the enrichment of 5hmC is a unique feature gained during PGC specification. This enrichment is likely due to highly active TET proteins as well as the lack of DNA replication in early pPGCs (Kobayashi et al., 2017). TET1, one of the three TET proteins, is moderately but not differentially expressed in pPGCs and gonadal hPGCs when compared to somatic lineages, suggesting the enhancement of TET1 activity may be posttranscriptional. Also, TET genes are expressed only at low level in pre-gonadal cyPGCs. Therefore, the enrichment of 5hmC in early cyPGCs will need further confirmation. The activity of TET proteins can be affected by metabolites (Yang et al., 2014, DiTroia et al., 2019), post-translational modification (Ciccarone et al., 2015) and alternative splicing (Zhang et al., 2016c). The upregulation of Vitamin C transporter SLC2A3 and DNA repair components PARP1 are also observed in pPGCs, which may enhance the activity of TET1 (DiTroia et al., 2019, Ciccarone et al., 2015). Other than enhancing the catalytic activity of TET1, it is also possible that the recruitment and accessible locus of TET1 may be

temporarily increased in early pPGCs when DNMT proteins are repressed (Charlton et al., 2020).

TET proteins and/or 5hmC have been shown to participate in multiple biological processes including cell cycle regulation (Prikrylova et al., 2019), cell differentiation (Verma et al., 2018, Rodríguez-Aguilera et al., 2020) and also reprogramming to pluripotency (Hu et al., 2014), suggesting TET activity may also be crucial for pPGC specification and maintaining expression of pluripotency genes. By targeting regulatory components such as bivalent promoters (Verma et al., 2018) and enhancers (Charlton et al., 2020) as well as modulating chromatin architecture such as promoter-enhancer looping (Fang et al., 2019), TET proteins and/or 5hmC may contribute to a permissive environment for PGC specification and also regulate transcription of key specifiers such as SOX17. Indeed, a recent study indicates supplementation of Vitamin C promotes hPGCLCs induction as well as upregulation of TET enzymes, 5hmC level and expression of key genes in hPGCLCs (Li et al., 2019).

It is recently reported that 5hmC can mark replication origin and its level is inversely correlated with replication time (Prikrylova et al., 2019), suggesting that the enrichment of 5hmC and lack of DNA replication may form a positive feedback loop and therefore active mechanisms of 5hmC removal may be crucial. Although upregulations of multiple BER pathway components such as UNG and PARP1 are observed in pPGCs, cyPGCs and hPGCs, the detailed roles of BER pathway in DNA demethylation in PGCs, for example if and which DNA glycosylates can directly excise 5hmC and initiate BER pathway, requires further investigation.

4.2.2 Absence of macroH2A1 throughout pPGC development and the early enrichment of H3K27me3 in pre-gonadal pPGCs

As mentioned in section Chapter 11.21.2.21.2.2.5, macroH2A is a conserved histone variant in vertebrates that involves in transcriptional regulation and has been linked to specific biological processes including DNA damage response, mitochondrial respiration and barriers to induced pluripotency (Pasque et al., 2012, Gaspar-Maia et al., 2013, Pasque et al., 2011, Barrero et al., 2013a).The

other intriguing epigenetic feature observed is the absence of macroH2A1, encoded by *H2AFY*, throughout the development of pPGCs. Downregulation of *H2AFY* is also evident in early cyPGCs, although it is moderately expressed in gonadal cyPGCs and hPGCs. Intriguingly, a homologues protein, macroH2A2, is absent in gonadal hPGC potentially via posttranscriptional regulation (Tang et al., 2015). These hint that the downregulation or absence of macroH2A may be a conserved feature in non-rodent PGCs, which will need further validation with in vitro models and early Cy monkey embryos.

The regulation and deposition of macroH2A are largely unclear. As E11 pig epiblast and the surrounding cells of the pre-migratory PGC cluster also lack macroH2A1, it is tempting to think that the largely absent macroH2A1 in pPGCs and potentially their precursors are a feature inherited from the pluripotent cells. As macroH2A has been suggest as a barrier against reprogramming to pluripotency and an important component of dense heterochromatin (Douet et al., 2017), H2AFY may be repressed transcriptionally in pPGCs to create a permissive chromatin status for the continuous expression of pluripotency genes and also the extensive epigenetic reprogramming. It should be noted that *H2AFY2*, which code macroH2A2, is also downregulated in pPGCs but not completely absent. It would be interesting to investigate if knockout or knockdown of H2AFY and/or H2AFY2 may increase the efficiency of PGCLC induction. Additionally, the function of an isoform of macroH2A1, macroH2A1.1, has been linked with PARP1 activity in metabolism and DNA repair mechanism. It is possible that the lack of macroH2A1 contributes to the increased activity of PARP1, which may play a role in the TET-mediated DNA demethylation and the metabolism shift (Hurtado-Bagès et al., 2020, Murata et al., 2019).

The dynamics of another histone modification, H3K27me3, is shown to be similar between pPGCs and hPGCs, but different to mPGCs after gonadal stage. An enrichment of H3K27me3 is observed in migratory PGCs of three species. In mPGC, this enrichment of H3K27me3 is shown to persist in gonadal stage (Seki et al., 2005, Seki et al., 2007, Prokopuk et al., 2017), despite of a transient loss around E11.5 reported by one study (Hajkova et al., 2008). In hPGC and pPGC, H3K27me3 is largely downregulated from Wk7 (Tang et al., 2015) and E36, respectively. On the other hand, in migratory

hPGCs and gonadal mPGCs, H3K27me3 show enrichment on the nuclear envelope (Prokopuk et al., 2017, Gomes Fernandes et al., 2018), which is not observed in pPGC in this project. It is possible that the association between H3K27me3 and nuclear envelope in pPGC is transient (e.g. between E17 and E25) and not captured in this project. The upregulation of H3K27me3 and its potential association with nuclear lamina may have a role in maintaining heterochromatin structure during the global DNA demethylation, removal of H3K9me2 and cell migration (Guelen et al., 2008, Harr et al., 2015), which may also contribute to nuclear rigidity (Gerlitz, 2020) and TE repression (Walter et al., 2016)

The mechanisms of this H3K27me3 dynamics in PGCs remains largely unknown. Prdm14 is suggested essential for the enrichment of H3K27me3 and the removal of H3K27me3 from Xi in migrating mPGCs (Mallol et al., 2019). However, as PRDM14 is almost absent in early pPGCs and possibly less critical for epigenetic reprogramming in hPGCs (Sybirna et al., 2020), the upregulation of H3K27me3 may depend on other mechanisms. Three key components of PRC2 complex, EZH1/2, SUZ12, and EED are present but not differentially expressed in pPGCs of all examined stages, but a few binding factors of PRC2 complex, such as PHF19 and RBBP7, are upregulated in pre-migratory pPGCs which may contribute to increased activity of PRC2 complex and hence the global upregulation of H3K27me3. As those PRC2 components are moderately expressed in gonadal pPGCs and no known H3K27me3 demethylase show upregulation, regulation of the following H3K27me3 removal may very likely to be posttranscriptional.

4.2.3 Early onset of XCR in pPGCs

In contrast to the XCR kinetics in mPGCs, an early and rapid XCR is detected in pre- and early-migratory pPGCs (E14-E17), shown by absence of the H3K27me3-marked Xi and detection of biallelic expression of X-linked genes. XCR continues in gonadal pPGCs (E31) but does not show an accelerated X-linked gene derepression as suggested in mPGCs. The extent of XCR also appear to be highly asynchronous among individual cells and genes (Table 2-1). Despite of evidences of an early XCR in pPGCs at both epigenetic and

transcriptional level, detailed kinetics and mechanisms of this event are still largely unknown.

First of all, the behaviour and function of Xist on Xi in pig germline and somatic cells require further characterization, and absence of Xist-coating may be required for confirmation of XCR initiation. Xist expression, shown by RT-PCR, is absent in 6 out of 21 (28.6%) E7.75 mPGCs, which is different from the absence of Xist-coating, shown by FISH, in only 4.7% of E7.75 mPGCs. This indicates that the absence of Xist transcript may not be sufficient as a marker for the initiation of XCR. Also, Xist expression is detected in human male germ cells and majority female germ cells from PGC stage to meiotic stage regardless of the X chromosome status (Li et al., 2017b, Vértesy et al., 2018). On the other hand, XACT, the antagonist of Xist in human preimplantation embryo, hasn't been detected in hPGCs. These suggest the activity and function of Xist and other lncRNA during XCR may vary largely between developmental stages and poorly conserved across species.

Consistent with the detection of mono-allelically-expressed genes even at early meiotic stage of female germ cell in both mouse (E14.5) and human (Wk14.4-17) (Vértesy et al., 2018), E31 pPGC still contains a substantial amount of monoallelically-expressed X-linked genes, despite of the lack of DNA methylation and repressive modification H3K27me3. It is not surprising that complete XCR in germ cells requires a prolonged period, as most of the known gene-silencing mechanisms, including epigenetic regulation, binding of silencing factors and remodelling of chromosome topological conformation (Chen et al., 2016a), have been reported to cooperate and contribute redundantly to the random and heritable XCI. Synergistic disruption of DNA methylation, topoisomerase activity and Xist-interacting protein still fail to achieve complete reactivation of all silenced X-linked genes in somatic cells (Minajigi et al., 2015). Further insights in the mechanisms of XCR in mammalian PGCs may shed light on potential treatment of X-linked diseases via reactivation of Xi. Behaviour of pluripotency factor (Augui et al., 2011, Payer et al., 2013), epigenetic modifications and their modifiers (Dixon-McDougall and Brown, 2015, Borensztein et al., 2017a), as well as Xist-binding architectural proteins (Minajigi et al., 2015) may all change dynamically during the transition of Xi to active state. Unsurprisingly our

single-cell transcriptome data of female pPGCs did not show differential expression for most Xist interactors, suggesting changes is likely to occur beyond transcriptional levels and vary between X-linked genes (da Rocha and Heard, 2017, Mira-Bontenbal and Gribnau, 2016, Cerase et al., 2015, Patil et al., 2016). It is worth noting that although TET proteins and 5hmC were suggested dispensable for the XCR during reprogramming to pluripotency (Pasque et al., 2014), I cannot rule out the possibility that the early enrichment of 5hmC may contribute to the early onset of XCR in pPGCs. Investigating the interactome and epigenetic status of reactivating X chromosome at locus-specific level, in both in vivo (pPGCs) and in vitro systems (hPGCLCs), may reveal further details about XCR mechanisms in PGCs.

4.2.4 Increased transcription of TE in pPGCs

DNA methylation has been suggested as the major host defence mechanism against the TE activity (Yoder et al., 1997), which can alter gene functions and cause genome instability (Molaro and Malik, 2016, Deniz et al., 2019). During the global DNA demethylation, the transcriptional repression of TE may largely depend on other epigenetic mechanisms including H3K9me3 and H3K27me3 (Liu et al., 2014, Walter et al., 2016). The increased transcription of TEs in pPGCs, evolutionarily young and old, during the global DNA demethylation was unexpected, as the expression of majority of TE families in gonadal hPGCs are downregulated compared to the surrounding somatic cells (Tang et al., 2015). This might be due to the difference between the less differentiated, multipotent E14 somatic cells in our project and the advanced human gonadal somatic cells. It is worth noting PGCs in pig, human and mouse all show higher expression of TEs when compared to pluripotent cells, namely E11 pig epiblast, human ICM (Guo et al., 2015) and E6.5 mouse epiblast (Ohno et al., 2013). Also, multiple post-transcriptional repression mechanisms for TE-activity, including the piRNA pathway, are activated in pPGCs, suggesting the mobilization of TE is likely to remain controlled (Reznik et al., 2019). Overall, although an increased expression of TE is observed in pPGCs, the activity of those potentially deleterious elements remains to be examined.

Despite of being potentially deleterious, studies have also acknowledged TE being a potent source of regulatory sequence and involving in multiple biological process (Diehl et al., 2020, Zhou et al., 2020). The upregulated expression of some TEs, such as L1, may have a regulatory role in PGC development as previously reported in neuronal precursors (Muotri et al., 2005, Jönsson et al., 2019), pre-implantation mouse embryos (Percharde et al., 2018) and mESCs (Lu et al., 2020)

4.3 Escapees and transgenerational epigenetic inheritance

Compared to previously reported in human and mouse, a smaller number of escapees have been identified in pPGCs, mostly overlapping transposon elements (TE-rich) and some occurring at single-copy genes (TE-poor). It appears that in pig, the global DNA demethylation during both waves of epigenetic reprogramming, in early embryo and PGCs, are more thorough compared to other mammalian species including human, mouse and cow (Ivanova et al., 2020). This may be partially due to the proportion of repetitive elements in the genome, with pig showing the least (44.57 %) followed by mouse (45.03 %), cow (49.38 %) and human (52.58 %) (RepeatMasker Genomic Datasets, 2020).

Nevertheless, the identified escapees in pPGC show conserved features with human and mouse, including association with metabolic and mental disorders. The existence of demethylation resistant loci suggests residual DNA methylation machineries still operate in gonadal PGCs. Recent studies in gonadal mPGC and also preimplantation embryo proposed that de novo methylation occurs during global DNA demethylation, which is limited by TET proteins (Hill et al., 2018, Hargan-Calvopina et al., 2016, Amouroux et al., 2016). As DNMT3A and DNMT3B are largely absent in gonadal hPGCs (Tang et al., 2015), the potency and frequency of de novo methylation will need further confirmation. It should be noted that the PBAT-seq used in the project, like all BS-seq, cannot distinguish 5mC from 5hmC. However, based on the LC-MS/MS results, only a small portion, if any, of the escapees will be hydroxymethylated. Tang et al. has proposed that highly expressed KRAB-

ZNF family in gonadal hPGC, binding of which leads to spreading of H3K9me3 (Groner et al., 2010), may play a role in recruiting the DNA methylation machinery to escapees. A recent study in preimplantation embryo also suggests H3K9me3 as a promoting and/or protection mechanism for de novo methylation on paternal genome (Albert et al., 2020). However, the KRAB-ZNF genes specifically expressed in gonadal hPGCs are mostly specific to primates and have no high-confidence orthologous identified in mouse or pig, suggesting this mechanism, if proven true, may be species-specific and therefore is more likely to target poorly conserved features such as TEs (Yang et al., 2017b, Wolf et al., 2020). Also, in E14.5 mouse male germ cell, tandem-arrayed transgenes are shown to be much less efficiently demethylated compared to the single-copy one, suggesting the demethylation efficiency varies depending on the genomic context (Newkirk et al., 2017). It remains to be explored if and how the residual DNA methylation machinery is actively recruited to escapees, especially at single-copy genes which shows greater conservation across species.

As the escapees show conserved and also unique association with phenotypes and disorders, another important question would be whether they response to environmental variations and have the potential to pass information to the following generations. Although intergenerational and transgenerational effects of altered environmental conditions (Kaspar et al., 2020, Perez and Lehner, 2019), have been observed, to date only a few primary, metastable epialleles has been identified in mammals (Bertozzi and Ferguson-Smith, 2020, Horsthemke, 2018). Studies have demonstrated that DNA methylation in mammalian PGC changes under environmental stress including deficiency of vitamin C (DiTroia et al., 2019) and endocrine disruptors (Beck et al., 2017, Skinner et al., 2019), suggesting escapees in PGCs can be environmental responsive. However, limited overlaps were seen between DMRs identified in different stages of germline development or in different generations (Beck et al., 2017, Skinner et al., 2019). Therefore, it remains to be examined if escapees, when altered in response to environmental changes, can evade both waves of epigenetic reprogramming and carry the epigenetic information across generations. On the other hand, changes in escapees may lead to alteration of local chromatin states including transcriptional factor binding

(Kremsky and Corces, 2020) and chromatin accessibility which may affect following epigenetic events such as DNA re-methylation. Therefore, changed escapees may function as inducers of a cascade of epigenetic alterations and the maintenance of transgenerational effect may rely on other mechanisms, which leads to different but related phenotypes observed in multiple generations. Additionally, it would be interesting to examine if TE activity and insertion rate is increased when escapees are fully erased. Increased TE activity may lead to changes in genomic content and if not lethal, may lead to altered phenotypes in next generations.

4.4 Summary

This project advances insight into the mechanism of pPGCs specification, and their subsequent development. The unique epigenetic signatures of newly specific pPGCs not only indicates an early initiation of epigenetic reprogramming, but also suggests a close link with pPGC specification. The unique epigenetic state, especially the enrichment of 5hmC and absence of macroH2A, may create a permissive state for pPGC specification. This epigenetic setting of pPGC may also interplay with other unique events observed, including the activation of DNA repair mechanism and the metabolic changes which can be crucial for pPGCs to maintain genome stability and cellular homeostasis. As pPGC going through the extensive epigenetic reprogramming, the global erasure of 5mC leads to another intriguing event, the transcriptional upregulation of retrotransposon, but whether the transposition activity is elevated remains to be explored in pPGCs. It will be of interest to study how in PGCs the activity of TEs is controlled without DNA methylation as the major repressive mechanism, as well as whether this unique period in germline provides the opportunity for accessing the regulatory potentials of TEs. Finally, when pPGCs reach basal methylation level, the escapees that evade global DNA demethylation are identified, changes of which may have the potential to initiate transgenerational effects. Still, the mechanisms and biological consequences for retaining escapees remain to be explored.

Investigations on very early hPGCs are exceptional (Tyser et al., 2020), especially during the assumed period of specification (~Wk2-3) when they are essentially inaccessible. Porcine embryos, which are relatively accessible compared to primates and ethically less challenging, have been used as the model to confirm the mechanism of non-rodent PGC specification which was observed in human in vitro models (Kobayashi et al., 2017). Based on the observation with in vitro models and early cyPGCs, many of the aforementioned events during pPGC specification, including the enrichment of 5hmC and absence of macroH2A1, may also occur during hPGC specification and can be further validated with the microfluidic model. As genetic editing and transgenerational studies are possible with pig, how epigenetic reprogramming may be affected, including the retain of escapees, by different environmental variations can be assessed. These suggest the studies of germline in pig and human will be mutually informative. Overall, this project provides a useful resource for further investigations on non-rodent germline that will increase comprehension of the underlying developmental mechanisms.

Reference

ADELMAN, C. A., LOLO, R. L., BIRKBAK, N. J., MURINA, O., MATSUZAKI, K., HOREJSI, Z., PARMAR, K., BOREL, V., SKEHEL, J. M. & STAMP, G. 2013. HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis. *Nature,* 502**,** 381-384.

AGOULNIK, A. I., LU, B., ZHU, Q., TRUONG, C., TY, M. T., ARANGO, N., CHADA, K. K. & BISHOP, C. E. 2002. A novel gene, Pog, is necessary for primordial germ cell proliferation in the mouse and underlies the germ cell deficient mutation, gcd. *Human molecular genetics,* 11**,** 3047-3053.

ALBERT, J. R., YEUNG, W. K. A., TORIYAMA, K., KOBAYASHI, H., HIRASAWA, R., BRIND'AMOUR, J., BOGUTZ, A., SASAKI, H. & LORINCZ, M. 2020. Maternal DNMT3A-dependent de novo methylation of the zygotic paternal genome inhibits gene expression in the early embryo. *bioRxiv***,** 2020.03.26.009977.

ALI, A. T., BOEHME, L., CARBAJOSA, G., SEITAN, V. C., SMALL, K. S. & HODGKINSON, A. 2019. Nuclear genetic regulation of the human mitochondrial transcriptome. *Elife,* 8**,** e41927.

AMOUROUX, R., NASHUN, B., SHIRANE, K., NAKAGAWA, S., HILL, P. W. S., D/'SOUZA, Z., NAKAYAMA, M., MATSUDA, M., TURP, A., NDJETEHE, E., ENCHEVA, V., KUDO, N. R., KOSEKI, H.,

SASAKI, H. & HAJKOVA, P. 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol,* 18**,** 225-233.

ARAMAKI, S., HAYASHI, K., KURIMOTO, K., OHTA, H., YABUTA, Y., IWANARI, H., MOCHIZUKI, Y., HAMAKUBO, T., KATO, Y., SHIRAHIGE, K. & SAITOU, M. 2013. A Mesodermal Factor, T, Specifies Mouse Germ Cell Fate by Directly Activating Germline Determinants. *Developmental Cell,* 27**,** 516-529.

ARAND, J., SPIELER, D., KARIUS, T., BRANCO, M. R., MEILINGER, D., MEISSNER, A., JENUWEIN, T., XU, G., LEONHARDT, H., WOLF, V. & WALTER, J. 2012. In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLOS Genetics,* 8**,** e1002750.

ARAND, J., WOSSIDLO, M., LEPIKHOV, K., PEAT, J. R., REIK, W. & WALTER, J. 2015. Selective impairment of methylation maintenance is the major cause of DNA methylation reprogramming in the early embryo. *Epigenetics & Chromatin,* 8**,** 1.

ARJAN-ODEDRA, S., SWANSON, C. M., SHERER, N. M., WOLINSKY, S. M. & MALIM, M. H. 2012. Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. *Retrovirology,* 9**,** 1-13.

ATCHISON, F. W., CAPEL, B. & MEANS, A. R. 2003. Pin1 regulates the timing of mammalian primordial germ cell proliferation. *Development,* 130**,** 3579-3586.

AUGUI, S., NORA, E. P. & HEARD, E. 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Reviews Genetics,* 12**,** 429.

BALATON, B. P. & BROWN, C. J. 2016. Escape Artists of the X Chromosome. *Trends in Genetics,* 32**,** 348-359.

BALLARÉ, C., LANGE, M., LAPINAITE, A., MARTIN, G. M., MOREY, L., PASCUAL, G., LIEFKE, R., SIMON, B., SHI, Y., GOZANI, O., CARLOMAGNO, T., BENITAH, S. A. & DI CROCE, L. 2012. Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. *Nature structural & molecular biology,* 19**,** 1257-1265.

BANASZYNSKI, L. A., ALLIS, C. D. & LEWIS, P. W. 2010. Histone variants in metazoan development. *Developmental cell,* 19**,** 662-674.

BARAN-GALE, J., CHANDRA, T. & KIRSCHNER, K. 2017. Experimental design for single-cell RNA sequencing. *Briefings in Functional Genomics,* 17**,** 233-239.

BARAU, J., TEISSANDIER, A., ZAMUDIO, N., ROY, S., NALESSO, V., HÉRAULT, Y., GUILLOU, F. & BOURC'HIS, D. 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science,* 354**,** 909-912.

BARRERO, MARÍA J., SESE, B., KUEBLER, B., BILIC, J., BOUE, S., MARTÍ, M. & IZPISUA BELMONTE, JUAN C. 2013a. Macrohistone Variants Preserve Cell Identity by Preventing the Gain of H3K4me2 during Reprogramming to Pluripotency. *Cell Reports,* 3**,** 1005-1011.

BARRERO, M. J., SESE, B., MARTÍ, M. & BELMONTE, J. C. I. 2013b. Macro histone variants are critical for the differentiation of human pluripotent cells. *Journal of Biological Chemistry,* 288**,** 16110-16116.

BARTKE, T., VERMEULEN, M., XHEMALCE, B., ROBSON, S. C., MANN, M. & KOUZARIDES, T. 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell,* 143**,** 470-484.

BASSOLS, A., COSTA, C., ECKERSALL, P. D., OSADA, J., SABRIÀ, J. & TIBAU, J. 2014. The pig as an animal model for human pathologies: A proteomics perspective. *PROTEOMICS – Clinical Applications,* 8**,** 715-731.

BAUSCH-FLUCK, D., HOFMANN, A., BOCK, T., FREI, A. P., CERCIELLO, F., JACOBS, A., MOEST, H., OMASITS, U., GUNDRY, R. L. & YOON, C. 2015. A mass spectrometric-derived cell surface protein atlas. *PloS one,* 10**,** e0121314.

BECK, D., SADLER-RIGGLEMAN, I. & SKINNER, M. K. 2017. Generational comparisons (F1 versus F3) of vinclozolin induced epigenetic transgenerational inheritance of sperm differential DNA methylation regions (epimutations) using MeDIP-Seq. *Environmental Epigenetics,* 3.

BECKER, J. S., NICETTO, D. & ZARET, K. S. 2016. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends in Genetics,* 32**,** 29-41.

BELL, E., CURRY, E. W., MEGCHELENBRINK, W., JOUNEAU, L., BROCHARD, V., TOMAZ, R. A., MAU, K. H. T., ATLASI, Y., DE SOUZA, R. A., MARKS, H., STUNNENBERG, H. G., JOUNEAU, A. & AZUARA, V. 2020. Dynamic CpG methylation delineates subregions within super-enhancers selectively decommissioned at the exit from naive pluripotency. *Nature Communications,* 11**,** 1112.

BERLETCH, J. B., MA, W., YANG, F., SHENDURE, J., NOBLE, W. S., DISTECHE, C. M. & DENG, X. 2015. Escape from X Inactivation Varies in Mouse Tissues. *PLoS Genetics,* 11**,** e1005079.

BERNEY, D. M., LOOIJENGA, L. H. J., IDREES, M., OOSTERHUIS, J. W., RAJPERT-DE MEYTS, E., ULBRIGHT, T. M. & SKAKKEBAEK, N. E. 2016. Germ cell neoplasia in situ (GCNIS): evolution of the current nomenclature for testicular pre-invasive germ cell malignancy. *Histopathology,* 69**,** 7-10.

BERNSTEIN, B. E., MIKKELSEN, T. S., XIE, X., KAMAL, M., HUEBERT, D. J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K., JAENISCH, R., WAGSCHAL, A., FEIL, R., SCHREIBER, S. L. & LANDER, E. S. 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell,* 125**,** 315-326.

BERTOZZI, T. M. & FERGUSON-SMITH, A. C. 2020. Metastable epialleles and their contribution to epigenetic inheritance in mammals. *Seminars in Cell & Developmental Biology,* 97**,** 93-105.

BESSLER, J. B., ANDERSEN, E. C. & VILLENEUVE, A. M. 2010. Differential Localization and Independent Acquisition of the H3K9me2 and H3K9me3 Chromatin Modifications in the Caenorhabditis elegans Adult Germ Line. *PLoS Genetics,* 6**,** e1000830.

BHANDARE, R., SCHUG, J., LE LAY, J., FOX, A., SMIRNOVA, O., LIU, C., NAJI, A. & KAESTNER, K. H. 2010. Genome-wide analysis of histone modifications in human pancreatic islets. *Genome Research,* 20**,** 428-433.

BHIN, J., JEONG, H.-S., KIM, J. S., SHIN, J. O., HONG, K. S., JUNG, H.-S., KIM, C., HWANG, D. & KIM, K.-S. 2015. PGC-Enriched miRNAs Control Germ Cell Development. *Molecules and Cells,* 38**,** 895-903.

BIELANSKA-OSUCHOWSKA, Z. 2006. Oogenesis in pig ovaries during the prenatal period: ultrastructure and morphometry. *Reprod Biol,* 6**,** 161-193.

BLAKE, G. E. T. & WATSON, E. D. 2016. Unravelling the complex mechanisms of transgenerational epigenetic inheritance. *Current Opinion in Chemical Biology,* 33**,** 101-107.

BLOOM, J. C., LOEHR, A. R., SCHIMENTI, J. C. & WEISS, R. S. 2019. Germline genome protection: implications for gamete quality and germ cell tumorigenesis. *Andrology,* 7**,** 516-526.

BOCK, C., TOMAZOU, E. M., BRINKMAN, A. B., MULLER, F., SIMMER, F., GU, H., JAGER, N., GNIRKE, A., STUNNENBERG, H. G. & MEISSNER, A. 2010. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotech,* 28**,** 1106-1114.

BOHACEK, J., VON WERDT, S. & MANSUY, I. M. 2016. Probing the germline-dependence of epigenetic inheritance using artificial insemination in mice. *Environmental Epigenetics,* 2**,** dvv015-dvv015.

BOKHARI, B. & SHARMA, S. 2019. Stress Marks on the Genome: Use or Lose? *International journal of molecular sciences,* 20**,** 364.

BORENSZTEIN, M., OKAMOTO, I., SYX, L., GUILBAUD, G., PICARD, C., ANCELIN, K., GALUPA, R., DIABANGOUAYA, P., SERVANT, N., BARILLOT, E., SURANI, A., SAITOU, M., CHEN, C.-J., ANASTASSIADIS, K. & HEARD, E. 2017a. Contribution of epigenetic landscapes and transcription factors to X-chromosome reactivation in the inner cell mass. *Nature Communications,* 8**,** 1297.

BORENSZTEIN, M., SYX, L., ANCELIN, K., DIABANGOUAYA, P., PICARD, C., LIU, T., LIANG, J.-B., VASSILEV, I., GALUPA, R., SERVANT, N., BARILLOT, E., SURANI, A., CHEN, C.-J. & HEARD, E. 2017b. Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nature structural & molecular biology,* 24**,** 226-233.

BOULARD, M., STORCK, S., CONG, R., PINTO, R., DELAGE, H. & BOUVET, P. 2010. Histone variant macroH2A1 deletion in mice causes female-specific steatosis. *Epigenetics & Chromatin,* 3**,** 8.

BOURC'HIS, D. & BESTOR, T. H. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature,* 431**,** 96-99.

BRAMSWIG, N. C., EVERETT, L. J., SCHUG, J., DORRELL, C., LIU, C., LUO, Y., STREETER, P. R., NAJI, A., GROMPE, M. & KAESTNER, K. H. 2013. Epigenomic plasticity enables human pancreatic α to β cell reprogramming. *The Journal of Clinical Investigation,* 123**,** 1275-1284.

BRAUNSCHWEIG, M., JAGANNATHAN, V., GUTZWILLER, A. & BEE, G. 2012. Investigations on Transgenerational Epigenetic Response Down the Male Line in F2 Pigs. *PLOS ONE,* 7**,** e30583.

BREILING, A. & LYKO, F. 2015. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics & Chromatin,* 8**,** 24.

BRICKER, D. K., TAYLOR, E. B., SCHELL, J. C., ORSAK, T., BOUTRON, A., CHEN, Y.-C., COX, J. E., CARDON, C. M., VAN VRANKEN, J. G., DEPHOURE, N., REDIN, C., BOUDINA, S., GYGI, S. P., BRIVET, M., THUMMEL, C. S. & RUTTER, J. 2012. A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans. *Science (New York, N.Y.),* 337**,** 96-100.

BRIEN, G. L., GAMBERO, G., O'CONNELL, D. J., JERMAN, E., TURNER, S. A., EGAN, C. M., DUNNE, E. J., JURGENS, M. C., WYNNE, K. & PIAO, L. 2012. Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nature Structural and Molecular Biology,* 19**,** 1273.

BRIEÑO-ENRÍQUEZ, M. A., LARRIBA, E. & DEL MAZO, J. 2016. Endocrine disrupters, microRNAs, and primordial germ cells: a dangerous cocktail. *Fertility and Sterility,* 106**,** 871-879.

BRINKMAN, A. B., GU, H., BARTELS, S. J., ZHANG, Y., MATARESE, F., SIMMER, F., MARKS, H., BOCK, C., GNIRKE, A. & MEISSNER, A. 2012. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome research,* 22**,** 1128-1138.

BROWN, C. J., BALLABIO, A., RUPERT, J. L., LAFRENIERE, R. G., GROMPE, M., TONLORENZI, R. & WILLARD, H. F. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature,* 349**,** 38.

BRYKCZYNSKA, U., HISANO, M., ERKEK, S., RAMOS, L., OAKELEY, E. J., ROLOFF, T. C., BEISEL, C., SCHUBELER, D., STADLER, M. B. & PETERS, A. H. F. M. 2010. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol,* 17**,** 679-687.

BULUT-KARSLIOGLU, A., INTI, A., RAMIREZ, F., BARENBOIM, M., ONISHI-SEEBACHER, M., ARAND, J., GALÁN, C., WINTER, G. E., ENGIST, B. & GERLE, B. 2014. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Molecular cell,* 55**,** 277-290.

BURGESS, R. J. & ZHANG, Z. 2013. Histone chaperones in nucleosome assembly and human disease. *Nature structural & molecular biology,* 20**,** 14-22.

BUSCHBECK, M. & HAKE, S. B. 2017. Variants of core histones and their roles in cell fate decisions, development and cancer. *Nature reviews Molecular cell biology,* 18**,** 299.

BUSCHBECK, M., URIBESALGO, I., WIBOWO, I., RUÉ, P., MARTIN, D., GUTIERREZ, A., MOREY, L., GUIGÓ, R., LÓPEZ-SCHIER, H. & DI CROCE, L. 2009. The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nature Structural &Amp; Molecular Biology,* 16**,** 1074.

BUTLER, A., HOFFMAN, P., SMIBERT, P., PAPALEXI, E. & SATIJA, R. 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology,* 36**,** 411.

BYGREN, L. O., MULLER, P., BRODIN, D., KAATI, G. & KRAL, J. G. 2017. Paternal grandparental exposure to crop failure or surfeit during a

childhood slow growth period and epigenetic marks on third generational growth-, glucoregulatory and stress genes. *bioRxiv*.

BYSKOV, A., HØYER, P., BJÖRKMAN, N., MØRK, A., OLSEN, B. & GRINSTED, J. 1986. Ultrastructure of germ cells and adjacent somatic cells correlated to initiation of meiosis in the fetal pig. *Anatomy and embryology,* 175**,** 57-67.

CALABRESE, A. N. & RADFORD, S. E. 2018. Mass spectrometry-enabled structural biology of membrane proteins. *Methods,* 147**,** 187-205.

CAMPOLO, F., GORI, M., FAVARO, R., NICOLIS, S., PELLEGRINI, M., BOTTI, F., ROSSI, P., JANNINI, E. A. & DOLCI, S. 2013. Essential Role of Sox2 for the Establishment and Maintenance of the Germ Cell Line. *STEM CELLS,* 31**,** 1408-1421.

CANOVAS, S., IVANOVA, E., ROMAR, R., GARCÍA-MARTÍNEZ, S., SORIANO-ÚBEDA, C., GARCÍA-VÁZQUEZ, F. A., SAADEH, H., ANDREWS, S., KELSEY, G. & COY, P. 2017. DNA methylation and gene expression changes derived from assisted reproductive technologies can be decreased by reproductive fluids. *eLife,* 6**,** e23670.

CANTONE, I. 2017. Reversal of X chromosome inactivation: lessons from pluripotent reprogramming of mouse and human somatic cells. *Journal of Translational Genetics and Genomics,* 1**,** 1-14.

CANTONE, I. & FISHER, A. G. 2017. Human X chromosome inactivation and reactivation: implications for cell reprogramming and disease. *Phil. Trans. R. Soc. B,* 372**,** 20160358.

CANZIO, D., CHANG, E. Y., SHANKAR, S., KUCHENBECKER, K. M., SIMON, M. D., MADHANI, H. D., NARLIKAR, G. J. & SADY, B. A. 2011. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome bridging mechanism for heterochromatin assembly. *Molecular cell,* 41**,** 67-81.

CAO, Z., ZHOU, N., ZHANG, Y., ZHANG, Y., WU, R., LI, Y., ZHANG, Y. & LI, N. 2014. Dynamic reprogramming of 5-hydroxymethylcytosine during early porcine embryogenesis. *Theriogenology,* 81**,** 496-508.

CARONE, B. R., FAUQUIER, L., HABIB, N., SHEA, J. M., HART, C. E., LI, R., BOCK, C., LI, C., ZAMORE, P. D., MEISSNER, A., WENG, Z., HOFMANN, H. A., FRIEDMAN, N. & RANDO, O. J. 2010. Paternally-induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell,* 143**,** 1084-1096.

CARREL, L. & WILLARD, H. F. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature,* 434**,** 400.

CECCALDI, R., SARANGI, P. & D'ANDREA, A. D. 2016. The Fanconi anaemia pathway: new players and new functions. *Nature Reviews Molecular Cell Biology,* 17**,** 337-349.

CERASE, A., PINTACUDA, G., TATTERMUSCH, A. & AVNER, P. 2015. Xist localization and function: new insights from multiple levels. *Genome Biology,* 16**,** 166.

CHADWICK, B. P. & WILLARD, H. F. 2001. Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. *Human Molecular Genetics,* 10**,** 1101-1113.

CHAKRAVARTHY, S., GUNDIMELLA, S. K. Y., CARON, C., PERCHE, P.-Y., PEHRSON, J. R., KHOCHBIN, S. & LUGER, K. 2005.

Structural Characterization of the Histone Variant macroH2A. *Molecular and Cellular Biology,* 25**,** 7616-7624.

CHANDLER, V. L. 2007. Paramutation: From Maize to Mice. *Cell,* 128**,** 641-645.

CHANG, C.-C., MA, Y., JACOBS, S., TIAN, X. C., YANG, X. & RASMUSSEN, T. P. 2005. A maternal store of macroH2A is removed from pronuclei prior to onset of somatic macroH2A expression in preimplantation embryos. *Developmental Biology,* 278**,** 367-380.

CHANGOLKAR, L. N., COSTANZI, C., LEU, N. A., CHEN, D., MCLAUGHLIN, K. J. & PEHRSON, J. R. 2007. Developmental Changes in Histone macroH2A1-Mediated Gene Regulation. *Molecular and Cellular Biology,* 27**,** 2758-2764.

CHARLET, J., DUYMICH, C. E., LAY, F. D., MUNDBJERG, K., SØRENSEN, K. D., LIANG, G. & JONES, P. A. 2016. Bivalent regions of cytosine methylation and H3K27 acetylation suggest an active role for DNA methylation at enhancers. *Molecular cell,* 62**,** 422-431.

CHARLTON, J., JUNG, E. J., MATTEI, A. L., BAILLY, N., LIAO, J., MARTIN, E. J., GIESSELMANN, P., BRÄNDL, B., STAMENOVA, E. K., MÜLLER, F.-J., KISKINIS, E., GNIRKE, A., SMITH, Z. D. & MEISSNER, A. 2020. TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nature Genetics,* 52**,** 819-827.

CHASTAIN, L. G. & SARKAR, D. K. 2017. Alcohol effects on the epigenome in the germline: Role in the inheritance of alcohol-related pathology. *Alcohol,* 60**,** 53-66.

CHAUMEIL, J., LE BACCON, P., WUTZ, A. & HEARD, E. 2006. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes & Development,* 20**,** 2223-2237.

CHAWENGSAKSOPHAK, K., SVINGEN, T., NG, E. T., EPP, T., SPILLER, C. M., CLARK, C., COOPER, H. & KOOPMAN, P. 2012. Loss of Wnt5a Disrupts Primordial Germ Cell Migration and Male Sexual Development in Mice1. *Biology of Reproduction,* 86**,** 4, 1-12-4, 1-12.

CHÉDIN, F., LIEBER, M. R. & HSIEH, C.-L. 2002. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences,* 99**,** 16916-16921.

CHEN, C., WANG, W., WANG, X., SHEN, D., WANG, S., WANG, Y., GAO, B., WIMMERS, K., MAO, J., LI, K. & SONG, C. 2019a. Retrotransposons evolution and impact on lncRNA and protein coding genes in pigs. *Mobile DNA,* 10**,** 19.

CHEN, C.-K., BLANCO, M., JACKSON, C., AZNAURYAN, E., OLLIKAINEN, N., SURKA, C., CHOW, A., CERASE, A., MCDONEL, P. & GUTTMAN, M. 2016a. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science,* 354**,** 468-472.

CHEN, D. & CLARK, A. T. 2018. Mitochondrial DNA selection in human germ cells. *Nature Cell Biology,* 20**,** 118-120.

CHEN, D., GELL, J. J., TAO, Y., SOSA, E. & CLARK, A. T. 2017a. Modeling human infertility with pluripotent stem cells. *Stem Cell Research,* 21**,** 187-192.

CHEN, D., LIU, W., LUKIANCHIKOV, A., HANCOCK, G. V., ZIMMERMAN, J., LOWE, M. G., KIM, R., GALIC, Z., IRIE, N., SURANI, M. A., JACOBSEN, S. E. & CLARK, A. T. 2017b. Germline competency of human embryonic stem cells depends on eomesodermin†. *Biology of Reproduction,* 97**,** 850-861.

CHEN, D., SUN, N., HOU, L., KIM, R., FAITH, J., ASLANYAN, M., TAO, Y., ZHENG, Y., FU, J., LIU, W., KELLIS, M. & CLARK, A. 2019b. Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. *Cell Reports,* 29**,** 4568-4582.e5.

CHEN, K., ZHAO, BOXUAN S. & HE, C. 2016b. Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chemical Biology,* 23**,** 74-85.

CHEN, Q., CHEN, Y., BIAN, C., FUJIKI, R. & YU, X. 2013. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature,* 493**,** 561-564.

CHEN, Q., YAN, M., CAO, Z., LI, X., ZHANG, Y., SHI, J., FENG, G.-H., PENG, H., ZHANG, X., ZHANG, Y., QIAN, J., DUAN, E., ZHAI, Q. & ZHOU, Q. 2016c. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science,* 351**,** 397-400.

CHEN, Q., YAN, W. & DUAN, E. 2016d. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat Rev Genet,* 17**,** 733-743.

CHEN, T., UEDA, Y., DODGE, J. E., WANG, Z. & LI, E. 2003. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology,* 23**,** 5594-5605.

CHNG, Z., TEO, A., PEDERSEN, R. A. & VALLIER, L. 2010. SIP1 Mediates Cell-Fate Decisions between Neuroectoderm and Mesendoderm in Human Pluripotent Stem Cells. *Cell Stem Cell,* 6**,** 59-70.

CHOI, I., KIM, R., LIM, H.-W., KAESTNER, K. H. & WON, K.-J. 2014. 5-hydroxymethylcytosine represses the activity of enhancers in embryonic stem cells: a new epigenetic signature for gene regulation. *BMC Genomics,* 15**,** 670.

CHOO, J. H., KIM, J. D., CHUNG, J. H., STUBBS, L. & KIM, J. 2006. Allele-specific deposition of macroH2A1 in imprinting control regions. *Human Molecular Genetics,* 15**,** 717-724.

CHUVA DE SOUSA LOPES, S. M., HAYASHI, K., SHOVLIN, T. C., MIFSUD, W., SURANI, M. A. & MCLAREN††, A. 2008. X Chromosome Activity in Mouse XX Primordial Germ Cells. *PLoS Genetics,* 4**,** e30.

CIABRELLI, F., COMOGLIO, F., FELLOUS, S., BONEV, B., NINOVA, M., SZABO, Q., XUEREB, A., KLOPP, C., ARAVIN, A., PARO, R., BANTIGNIES, F. & CAVALLI, G. 2017. Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. *Nat Genet,* 49**,** 876-886.

CICCARONE, F., VALENTINI, E., ZAMPIERI, M. & CAIAFA, P. 2015. 5mC-hydroxylase activity is influenced by the PARylation of TET1 enzyme. *Oncotarget,* 6**,** 24333-24347.

CICCONE, D. N., SU, H., HEVI, S., GAY, F., LEI, H., BAJKO, J., XU, G., LI, E. & CHEN, T. 2009. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature,* 461**,** 415-418.

CIMMINO, L. & AIFANTIS, I. 2017. Alternative roles for oxidized mCs and TETs. *Current Opinion in Genetics & Development,* 42**,** 1-7.

CINGOLANI, P., PLATTS, A., WANG, L. L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly,* 6**,** 80-92.

CLARK, S. J., ARGELAGUET, R., KAPOURANI, C.-A., STUBBS, T. M., LEE, H. J., ALDA-CATALINAS, C., KRUEGER, F., SANGUINETTI, G., KELSEY, G., MARIONI, J. C., STEGLE, O. & REIK, W. 2018. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nature communications,* 9**,** 781-781.

CLARK, S. J., SMALLWOOD, S. A., LEE, H. J., KRUEGER, F., REIK, W. & KELSEY, G. 2017. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nature protocols,* 12**,** 534-547.

CONGRAS, A., YERLE-BOUISSOU, M., PINTON, A., VIGNOLES, F., LIAUBET, L., FERCHAUD, S. & ACLOQUE, H. 2014. Sperm DNA Methylation Analysis in Swine Reveals Conserved and Species-Specific Methylation Patterns and Highlights an Altered Methylation at the GNAS Locus in Infertile Boars. *Biology of Reproduction,* 91**,** 137, 1-14.

CONTRERAS CARO DEL CASTILLO, D. A. 2011. *Lineage segregation and primordial germ cell development in the pig / by David Alejandro Contreras Caro Del Castillo.* Thesis (PhD)--University of Nottingham.

CORTELLINO, S., XU, J., SANNAI, M., MOORE, R., CARETTI, E., CIGLIANO, A., LE COZ, M., DEVARAJAN, K., WESSELS, A., SOPRANO, D., ABRAMOWITZ, LARA K., BARTOLOMEI, MARISA S., RAMBOW, F., BASSI, MARIA R., BRUNO, T., FANCIULLI, M., RENNER, C., KLEIN-SZANTO, ANDRES J., MATSUMOTO, Y., KOBI, D., DAVIDSON, I., ALBERTI, C., LARUE, L. & BELLACOSA, A. 2011. Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell,* 146**,** 67-79.

COSTA, Y., DING, J., THEUNISSEN, T. W., FAIOLA, F., HORE, T. A., SHLIAHA, P. V., FIDALGO, M., SAUNDERS, A., LAWRENCE, M., DIETMANN, S., DAS, S., LEVASSEUR, D. N., LI, Z., XU, M., REIK, W., SILVA, J. C. R. & WANG, J. 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature,* 495**,** 370-374.

COSTANTINI, M., AULETTA, F. & BERNARDI, G. 2011. The Distributions of "New" and "Old" Alu Sequences in the Human Genome: The

Solution of a “Mystery”. *Molecular Biology and Evolution,* 29**,** 421-427.

COSTANZI, C. & PEHRSON, J. R. 1998. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature,* 393**,** 599-601.

COSTANZI, C. & PEHRSON, J. R. 2001. MACROH2A2, a New Member of the MACROH2A Core Histone Family. *Journal of Biological Chemistry,* 276**,** 21776-21784.

COSTANZI, C., STEIN, P., WORRAD, D. M., SCHULTZ, R. M. & PEHRSON, J. R. 2000. Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. *Development,* 127**,** 2283-2289.

COTTON, A. M., GE, B., LIGHT, N., ADOUE, V., PASTINEN, T. & BROWN, C. J. 2013. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biology,* 14**,** R122.

COURT, F., TAYAMA, C., ROMANELLI, V., MARTIN-TRUJILLO, A., IGLESIAS-PLATAS, I., OKAMURA, K., SUGAHARA, N., SIMÓN, C., MOORE, H., HARNESS, J. V., KEIRSTEAD, H., SANCHEZ-MUT, J. V., KANEKI, E., LAPUNZINA, P., SOEJIMA, H., WAKE, N., ESTELLER, M., OGATA, T., HATA, K., NAKABAYASHI, K. & MONK, D. 2014. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome research,* 24**,** 554-569.

CREPPE, C., JANICH, P., CANTARIÑO, N., NOGUERA, M., VALERO, V., MUSULÉN, E., DOUET, J., POSAVEC, M., MARTÍN-CABALLERO, J., SUMOY, L., DI CROCE, L., BENITAH, S. A. & BUSCHBECK, M. 2012. MacroH2A1 Regulates the Balance between Self-Renewal and Differentiation Commitment in Embryonic and Adult Stem Cells. *Molecular and Cellular Biology,* 32**,** 1442-1452.

D'SOUZA, A. R. & MINCZUK, M. 2018. Mitochondrial transcription and translation: overview. *Essays in biochemistry,* 62**,** 309-320.

DA ROCHA, S. T. & HEARD, E. 2017. Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nature Structural &Amp; Molecular Biology,* 24**,** 197.

DAHL, J. A., JUNG, I., AANES, H., GREGGAINS, G. D., MANAF, A., LERDRUP, M., LI, G., KUAN, S., LI, B., LEE, A. Y., PREISSL, S., JERMSTAD, I., HAUGEN, M. H., SUGANTHAN, R., BJØRÅS, M., HANSEN, K., DALEN, K. T., FEDORCSAK, P., REN, B. & KLUNGLAND, A. 2016. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature,* 537**,** 548-552.

DAWLATY, MEELAD M., BREILING, A., LE, T., RADDATZ, G., BARRASA, M. I., CHENG, ALBERT W., GAO, Q., POWELL, BENJAMIN E., LI, Z., XU, M., FAULL, KYM F., LYKO, F. & JAENISCH, R. 2013. Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Developmental Cell,* 24**,** 310-323.

DAXINGER, L. & WHITELAW, E. 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet,* 13**,** 153-162.

DE FAZIO, S., BARTONICEK, N., DI GIACOMO, M., ABREU-GOODGER, C., SANKAR, A., FUNAYA, C., ANTONY, C., MOREIRA, P. N., ENRIGHT, A. J. & O'CARROLL, D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature,* 480**,** 259-263.

DE FELICI, M. 2013. Origin, Migration, and Proliferation of Human Primordial Germ Cells. *In:* COTICCHIO, G., ALBERTINI, D. F. & DE SANTIS, L. (eds.) *Oogenesis.* Springer London.

DE MENDOZA, A., BONNET, A., VARGAS-LANDIN, D. B., JI, N., LI, H., YANG, F., LI, L., HORI, K., PFLUEGER, J., BUCKBERRY, S., OHTA, H., ROSIC, N., LESAGE, P., LIN, S. & LISTER, R. 2018. Recurrent acquisition of cytosine methyltransferases into eukaryotic retrotransposons. *Nature Communications,* 9**,** 1341.

DE NAPOLES, M., NESTEROVA, T. & 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. M. C., HAYASHI, K. & SURANI, M. A. 2007. Proximal visceral endoderm and extraembryonic ectoderm regulate the formation of primordial germ cell precursors. *BMC Developmental Biology,* 7**,** 140.

DELL'ORSO, S., WANG, A. H., SHIH, H.-Y., SASO, K., BERGHELLA, L., GUTIERREZ-CRUZ, G., LADURNER, ANDREAS G., O'SHEA, JOHN J., SARTORELLI, V. & ZARE, H. 2016. The Histone Variant MacroH2A1.2 Is Necessary for the Activation of Muscle Enhancers and Recruitment of the Transcription Factor Pbx1. *Cell Reports,* 14**,** 1156-1168.

DENG, X., BERLETCH, J. B., NGUYEN, D. K. & DISTECHE, C. M. 2014. X chromosome regulation: diverse patterns in development, tissues and disease. *Nature Reviews Genetics,* 15**,** nrg3687.

DENG, X., MA, W., RAMANI, V., HILL, A., YANG, F., AY, F., BERLETCH, J. B., BLAU, C. A., SHENDURE, J., DUAN, Z., NOBLE, W. S. & DISTECHE, C. M. 2015. Bipartite structure of the inactive mouse X chromosome. *Genome Biology,* 16**,** 152.

DENIZ, Ö., FROST, J. M. & BRANCO, M. R. 2019. Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics,* 20**,** 417-431.

DEPLUS, R., DELATTE, B., SCHWINN, M. K., DEFRANCE, M., MÉNDEZ, J., MURPHY, N., DAWSON, M. A., VOLKMAR, M., PUTMANS, P., CALONNE, E., SHIH, A. H., LEVINE, R. L., BERNARD, O., MERCHER, T., SOLARY, E., URH, M., DANIELS, D. L. & FUKS, F. 2013. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO Journal,* 32**,** 645-655.

DI CARLO, A., TRAVIA, G. & DE FELICI, M. 2002. The meiotic specific synaptonemal complex protein SCP3 is expressed by female and male primordial germ cells of the mouse embryo. *International Journal of Developmental Biology,* 44**,** 241-244.

DIEHL, A. G., OUYANG, N. & BOYLE, A. P. 2020. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nature Communications,* 11**,** 1796.

DITROIA, S. P., PERCHARDE, M., GUERQUIN, M.-J., WALL, E., COLLIGNON, E., EBATA, K. T., MESH, K., MAHESULA, S., AGATHOCLEOUS, M., LAIRD, D. J., LIVERA, G. & RAMALHO-SANTOS, M. 2019. Maternal vitamin C regulates reprogramming of DNA methylation and germline development. *Nature,* 573**,** 271-275.

DIXON-MCDOUGALL, T. & BROWN, C. 2015. The making of a Barr body: the mosaic of factors that eXIST on the mammalian inactive X chromosome. *Biochemistry and Cell Biology,* 94**,** 56-70.

DODGE, J. E., KANG, Y.-K., BEPPU, H., LEI, H. & LI, E. 2004. Histone H3-K9 Methyltransferase ESET Is Essential for Early Development. *Molecular and Cellular Biology,* 24**,** 2478-2486.

DOLATABADI, S., CANDIA, J., AKRAP, N., VANNAS, C., TESAN TOMIC, T., LOSERT, W., LANDBERG, G., ÅMAN, P. & STÅHLBERG, A. 2017. Cell cycle and cell size dependent gene expression reveals distinct subpopulations at single-cell level. *Frontiers in genetics,* 8**,** 1.

DONKIN, I., VERSTEYHE, S., INGERSLEV, L. R., QIAN, K., MECHTA, M., NORDKAP, L., MORTENSEN, B., APPEL, E. V. R., JØRGENSEN, N. & KRISTIANSEN, V. B. 2016. Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell metabolism,* 23**,** 369-378.

DOUET, J., CORUJO, D., MALINVERNI, R., RENAULD, J., SANSONI, V., POSAVEC MARJANOVIĆ, M., CANTARIÑO, N., VALERO, V., MONGELARD, F., BOUVET, P., IMHOF, A., THIRY, M. & BUSCHBECK, M. 2017. MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. *Journal of Cell Science,* 130**,** 1570-1582.

ECCO, G., IMBEAULT, M. & TRONO, D. 2017. KRAB zinc finger proteins. *Development,* 144**,** 2719-2729.

EGUIZABAL, C., HERRERA, L., DE OÑATE, L., MONTSERRAT, N., HAJKOVA, P. & IZPISUA BELMONTE, J. C. 2016. Characterization of the Epigenetic Changes During Human Gonadal Primordial Germ Cells Reprogramming. *STEM CELLS,* 34**,** 2418-2428.

ENGREITZ, J. M., PANDYA-JONES, A., MCDONEL, P., SHISHKIN, A., SIROKMAN, K., SURKA, C., KADRI, S., XING, J., GOREN, A., LANDER, E. S., PLATH, K. & GUTTMAN, M. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X-chromosome. *Science (New York, N.Y.),* 341**,** 1237973-1237973.

EPSTEIN, C. J. 1969. Mammalian oocytes: X chromosome activity. *Science,* 163**,** 1078-1079.

ESCAMILLA-DEL-ARENAL, M., DA ROCHA, S. T. & HEARD, E. 2011. Evolutionary diversity and developmental regulation of X-chromosome inactivation. *Human Genetics,* 130**,** 307-327.

FANG, S., LI, J., XIAO, Y., LEE, M., GUO, L., HAN, W., LI, T., HILL, M. C., HONG, T., MO, W., XU, R., ZHANG, P., WANG, F., CHANG, J., ZHOU, Y., SUN, D., MARTIN, J. F. & HUANG, Y. 2019. Tet

inactivation disrupts YY1 binding and long-range chromatin interactions during embryonic heart development. *Nature Communications,* 10**,** 4297.

FANG, X., MOU, Y., HUANG, Z., LI, Y., HAN, L., ZHANG, Y., FENG, Y., CHEN, Y., JIANG, X., ZHAO, W., SUN, X., XIONG, Z., YANG, L., LIU, H., FAN, D., MAO, L., REN, L., LIU, C., WANG, J., LI, K., WANG, G., YANG, S., LAI, L., ZHANG, G., LI, Y., WANG, J., BOLUND, L., YANG, H., WANG, J., FENG, S., LI, S. & DU, Y. 2012. The sequence and analysis of a Chinese pig genome. *GigaScience,* 1**,** 16-16.

FEENEY, A., NILSSON, E. & SKINNER, M. K. 2014. Epigenetics and transgenerational inheritance in domesticated farm animals. *Journal of animal science and biotechnology,* 5**,** 48.

FICZ, G., BRANCO, M. R., SEISENBERGER, S., SANTOS, F., KRUEGER, F., HORE, T. A., MARQUES, C. J., ANDREWS, S. & REIK, W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature,* 473**,** 398-402.

FLÉCHON, J.-E., DEGROUARD, J. & FLÉCHON, B. 2004. Gastrulation events in the prestreak pig embryo: Ultrastructure and cell markers. *genesis,* 38**,** 13-25.

FLOROS, V. I., PYLE, A., DIETMANN, S., WEI, W., TANG, W. C. W., IRIE, N., PAYNE, B., CAPALBO, A., NOLI, L., COXHEAD, J., HUDSON, G., CROSIER, M., STRAHL, H., KHALAF, Y., SAITOU, M., ILIC, D., SURANI, M. A. & CHINNERY, P. F. 2018. Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. *Nature Cell Biology,* 20**,** 144-151.

FRIED, V. 1992. MacroH2A, a core histone containing a large nonhistone region. *Science,* 257**,** 1398-1400.

FUCHSBERGER, C., FLANNICK, J., TESLOVICH, T. M., MAHAJAN, A., AGARWALA, V., GAULTON, K. J., MA, C., FONTANILLAS, P., MOUTSIANAS, L., MCCARTHY, D. J., RIVAS, M. A., PERRY, J. R. B., SIM, X., BLACKWELL, T. W., ROBERTSON, N. R., RAYNER, N. W., CINGOLANI, P., LOCKE, A. E., TAJES, J. F., HIGHLAND, H. M., DUPUIS, J., CHINES, P. S., LINDGREN, C. M., HARTL, C., JACKSON, A. U., CHEN, H., HUYGHE, J. R., VAN DE BUNT, M., PEARSON, R. D., KUMAR, A., MÜLLER-NURASYID, M., GRARUP, N., STRINGHAM, H. M., GAMAZON, E. R., LEE, J., CHEN, Y., SCOTT, R. A., BELOW, J. E., CHEN, P., HUANG, J., GO, M. J., STITZEL, M. L., PASKO, D., PARKER, S. C. J., VARGA, T. V., GREEN, T., BEER, N. L., DAY-WILLIAMS, A. G., FERREIRA, T., FINGERLIN, T., HORIKOSHI, M., HU, C., HUH, I., IKRAM, M. K., KIM, B.-J., KIM, Y., KIM, Y. J., KWON, M.-S., LEE, J., LEE, S., LIN, K.-H., MAXWELL, T. J., NAGAI, Y., WANG, X., WELCH, R. P., YOON, J., ZHANG, W., BARZILAI, N., VOIGHT, B. F., HAN, B.-G., JENKINSON, C. P., KUULASMAA, T., KUUSISTO, J., MANNING, A., NG, M. C. Y., PALMER, N. D., BALKAU, B., STANČÁKOVÁ, A., ABBOUD, H. E., BOEING, H., GIEDRAITIS, V., PRABHAKARAN, D., GOTTESMAN, O., SCOTT, J., CAREY, J., KWAN, P., GRANT, G., SMITH, J. D., NEALE, B. M., PURCELL, S., BUTTERWORTH, A. S., HOWSON, J. M. M., LEE, H. M., LU, Y.,

KWAK, S.-H., ZHAO, W., DANESH, J., LAM, V. K. L., PARK, K. S., SALEHEEN, D., et al. 2016. The genetic architecture of type 2 diabetes. *Nature,* 536**,** 41-47.

FUKUDA, A., TANINO, M., MATOBA, R., UMEZAWA, A. & AKUTSU, H. 2015. Imbalance between the expression dosages of X-chromosome and autosomal genes in mammalian oocytes. *Scientific Reports,* 5**,** 14101.

GAMBLE, M. J., FRIZZELL, K. M., YANG, C., KRISHNAKUMAR, R. & KRAUS, W. L. 2010. The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes & development,* 24**,** 21-32.

GAO, F., NIU, Y., SUN, Y. E., LU, H., CHEN, Y., LI, S., KANG, Y., LUO, Y., SI, C., YU, J., LI, C., SUN, N., SI, W., WANG, H., JI, W. & TAN, T. 2017. De novo DNA methylation during monkey pre-implantation embryogenesis. *Cell Res*.

GAO, F., XIA, Y., WANG, J., LUO, H., GAO, Z., HAN, X., ZHANG, J., HUANG, X., YAO, Y., LU, H., YI, N., ZHOU, B., LIN, Z., WEN, B., ZHANG, X., YANG, H. & WANG, J. 2013. Integrated detection of both 5-mC and 5-hmC by high-throughput tag sequencing technology highlights methylation reprogramming of bivalent genes during cellular differentiation. *Epigenetics,* 8**,** 421-430.

GASPAR-MAIA, A., QADEER, Z. A., HASSON, D., RATNAKUMAR, K., ADRIAN LEU, N., LEROY, G., LIU, S., COSTANZI, C., VALLE-GARCIA, D., SCHANIEL, C., LEMISCHKA, I., GARCIA, B., PEHRSON, J. R. & BERNSTEIN, E. 2013. MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. *Nature Communications,* 4**,** 1565.

GELL, J. J., ZHAO, J., CHEN, D., HUNT, T. J. & CLARK, A. T. 2018. PRDM14 is expressed in germ cell tumors with constitutive overexpression altering human germline differentiation and proliferation. *Stem Cell Research,* 27**,** 46-56.

GERLITZ, G. 2020. The Emerging Roles of Heterochromatin in Cell Migration. *Frontiers in cell and developmental biology,* 8**,** 394-394.

GILLESPIE, P. J., GAMBUS, A. & BLOW, J. J. 2012. Preparation and use of Xenopus egg extracts to study DNA replication and chromatin associated proteins. *Methods (San Diego, Calif.),* 57**,** 203-213.

GKOUNTELA, S., ZHANG, KELVIN X., SHAFIQ, TIASHA A., LIAO, W.-W., HARGAN-CALVOPIÑA, J., CHEN, P.-Y. & CLARK, AMANDER T. 2015. DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell,* 161**,** 1425-1436.

GOMES FERNANDES, M., BIALECKA, M., SALVATORI, D. C. F. & CHUVA DE SOUSA LOPES, S. M. 2018. Characterization of migratory primordial germ cells in the aorta-gonad-mesonephros of a 4.5-week-old human embryo: a toolbox to evaluate in vitro early gametogenesis. *Molecular Human Reproduction,* 24**,** 233-243.

GONZALEZ-BULNES, A., ASTIZ, S., OVILO, C., LOPEZ-BOTE, C. J., SANCHEZ-SANCHEZ, R., PEREZ-SOLANA, M. L., TORRES-ROVIRA, L., AYUSO, M. & GONZALEZ, J. 2014. Early-postnatal changes in adiposity and lipids profile by transgenerational

developmental programming in swine with obesity/leptin resistance. *Journal of Endocrinology,* 223**,** M17-M29.

GONZALEZ-MUNOZ, E., ARBOLEDA-ESTUDILLO, Y., CHANUMOLU, S. K., OTU, H. H. & CIBELLI, J. B. 2019. Zebrafish macroH2A variants have distinct embryo localization and function. *Scientific Reports,* 9**,** 8632.

GRAHAM, B., MARCAIS, A., DHARMALINGAM, G., CARROLL, T., KANELLOPOULOU, C., GRAUMANN, J., NESTEROVA, TATYANA B., BERMANGE, A., BRAZAUSKAS, P., XELLA, B., KRIAUCIONIS, S., HIGGS, DOUGLAS R., BROCKDORFF, N., MANN, M., FISHER, AMANDA G. & MERKENSCHLAGER, M. 2016. MicroRNAs of the miR-290–295 Family Maintain Bivalency in Mouse Embryonic Stem Cells. *Stem Cell Reports,* 6**,** 635-642.

GRAVES, J. A. M. 2015. Evolution of vertebrate sex chromosomes and dosage compensation. *Nature Reviews Genetics,* 17**,** 33.

GROENEN, M. A. M., ARCHIBALD, A. L., UENISHI, H., TUGGLE, C. K., TAKEUCHI, Y., ROTHSCHILD, M. F., ROGEL-GAILLARD, C., PARK, C., MILAN, D., MEGENS, H.-J., LI, S., LARKIN, D. M., KIM, H., FRANTZ, L. A. F., CACCAMO, M., AHN, H., AKEN, B. L., ANSELMO, A., ANTHON, C., AUVIL, L., BADAOUI, B., BEATTIE, C. W., BENDIXEN, C., BERMAN, D., BLECHA, F., BLOMBERG, J., BOLUND, L., BOSSE, M., BOTTI, S., BUJIE, Z., BYSTROM, M., CAPITANU, B., CARVALHO-SILVA, D., CHARDON, P., CHEN, C., CHENG, R., CHOI, S.-H., CHOW, W., CLARK, R. C., CLEE, C., CROOIJMANS, R. P. M. A., DAWSON, H. D., DEHAIS, P., DE SAPIO, F., DIBBITS, B., DROU, N., DU, Z.-Q., EVERSOLE, K., FADISTA, J., FAIRLEY, S., FARAUT, T., FAULKNER, G. J., FOWLER, K. E., FREDHOLM, M., FRITZ, E., GILBERT, J. G. R., GIUFFRA, E., GORODKIN, J., GRIFFIN, D. K., HARROW, J. L., HAYWARD, A., HOWE, K., HU, Z.-L., HUMPHRAY, S. J., HUNT, T., HORNSHØJ, H., JEON, J.-T., JERN, P., JONES, M., JURKA, J., KANAMORI, H., KAPETANOVIC, R., KIM, J., KIM, J.-H., KIM, K.-W., KIM, T.-H., LARSON, G., LEE, K., LEE, K.-T., LEGGETT, R., LEWIN, H. A., LI, Y., LIU, W., LOVELAND, J. E., LU, Y., LUNNEY, J. K., MA, J., MADSEN, O., MANN, K., MATTHEWS, L., MCLAREN, S., MOROZUMI, T., MURTAUGH, M. P., NARAYAN, J., TRUONG NGUYEN, D., NI, P., OH, S.-J., ONTERU, S., PANITZ, F., PARK, E.-W., et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature,* 491**,** 393.

GRONER, A. C., MEYLAN, S., CIUFFI, A., ZANGGER, N., AMBROSINI, G., DÉNERVAUD, N., BUCHER, P. & TRONO, D. 2010. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS genetics,* 6**,** e1000869-e1000869.

GU, T.-P., GUO, F., YANG, H., WU, H.-P., XU, G.-F., LIU, W., XIE, Z.-G., SHI, L., HE, X., JIN, S.-G., IQBAL, K., SHI, Y. G., DENG, Z., SZABO, P. E., PFEIFER, G. P., LI, J. & XU, G.-L. 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature,* 477**,** 606-610.

GUELEN, L., PAGIE, L., BRASSET, E., MEULEMAN, W., FAZA, M. B., TALHOUT, W., EUSSEN, B. H., DE KLEIN, A., WESSELS, L., DE LAAT, W. & VAN STEENSEL, B. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature,* 453**,** 948-951.

GUENATRI, M., DUFFIÉ, R., IRANZO, J., FAUQUE, P. & BOURC'HIS, D. 2013. Plasticity in Dnmt3L-dependent and -independent modes of de novo methylation in the developing mouse embryo. *Development,* 140**,** 562-572.

GUIBERT, S., FORNÉ, T. & WEBER, M. 2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Research,* 22**,** 633-641.

GUO, F., LI, X., LIANG, D., LI, T., ZHU, P., GUO, H., WU, X., WEN, L., GU, T.-P., HU, B., WALSH, COLUM P., LI, J., TANG, F. & XU, G.-L. 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., WANG, W., LI, R., YAN, J., ZHI, X., ZHANG, Y., JIN, H., ZHANG, W., HOU, Y., ZHU, P., LI, J., ZHANG, L., LIU, S., REN, Y., ZHU, X., WEN, L., GAO, Y. Q., TANG, F. & QIAO, J. 2015. The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell,* 161**,** 1437-1452.

GUO, H., CHITIPROLU, M., GAGNON, D., MENG, L., PEREZ-IRATXETA, C., LAGACE, D. & GIBBINGS, D. 2014b. Autophagy supports genomic stability by degrading retrotransposon RNA. *Nature communications,* 5**,** 1-11.

GUO, H., HU, B., YAN, L., YONG, J., WU, Y., GAO, Y., GUO, F., HOU, Y., FAN, X., DONG, J., WANG, X., ZHU, X., YAN, J., WEI, Y., JIN, H., ZHANG, W., WEN, L., TANG, F. & QIAO, J. 2016. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res*.

GUO, H., ZHU, P., YAN, L., LI, R., HU, B., LIAN, Y., YAN, J., REN, X., LIN, S., LI, J., JIN, X., SHI, X., LIU, P., WANG, X., WANG, W., WEI, Y., LI, X., GUO, F., WU, X., FAN, X., YONG, J., WEN, L., XIE, S. X., TANG, F. & QIAO, J. 2014c. The DNA methylation landscape of human early embryos. *Nature,* 511**,** 606-610.

HACKETT, J. A., HUANG, Y., GUNESDOGAN, U., HOLM-GRETARSSON, K., KOBAYASHI, T. & SURANI, A. M. 2018. Tracing the Transitions from Pluripotency to Germ Cell Fate with CRISPR Screening. *bioRxiv*.

HACKETT, J. A., REDDINGTON, J. P., NESTOR, C. E., DUNICAN, D. S., BRANCO, M. R., REICHMANN, J., REIK, W., SURANI, M. A., ADAMS, I. R. & MEEHAN, R. R. 2012a. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development,* 139**,** 3623-3632.

HACKETT, J. A., SENGUPTA, R., ZYLICZ, J. J., MURAKAMI, K., LEE, C., DOWN, T. A. & SURANI, M. A. 2013. Germline DNA Demethylation Dynamics and Imprint Erasure through 5-hydroxymethylcytosine. *Science (New York, N.Y.),* 339**,** 10.1126/science.1229277.

HACKETT, J. A., ZYLICZ, J. J. & SURANI, M. A. 2012b. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends in Genetics,* 28**,** 164-174.

HAHN, M. A., SZABÓ, P. E. & PFEIFER, G. P. 2014. 5-hydroxymethylcytosine: a stable or transient DNA modification? *Genomics,* 104**,** 314-323.

HAJKOVA, P., ANCELIN, K., WALDMANN, T., LACOSTE, N., LANGE, U. C., CESARI, F., LEE, C., ALMOUZNI, G., SCHNEIDER, R. & SURANI, M. A. 2008. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature,* 452.

HAJKOVA, P., JEFFRIES, S. J., LEE, C., MILLER, N., JACKSON, S. P. & SURANI, M. A. 2010. Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. *Science,* 329**,** 78-82.

HAMER, G. & DE ROOIJ, D. G. 2018. Mutations causing specific arrests in the development of mouse primordial germ cells and gonocytes. *Biology of Reproduction,* 99**,** 75-86.

HAMMOUD, SAHER S., LOW, DIANA H. P., YI, C., CARRELL, DOUGLAS T., GUCCIONE, E. & CAIRNS, BRADLEY R. 2014. Chromatin and Transcription Transitions of Mammalian Adult Germline Stem Cells and Spermatogenesis. *Cell Stem Cell,* 15**,** 239-253.

HANCOCK, G. V., LIU, W., PERETZ, L., CHEN, D., GELL, J. J., COLLIER, A. J., ZAMUDIO, J. R., PLATH, K. & CLARK, A. T. 2020. Divergent roles for KLF4 and TFCP2L1 in Naive Ground State Pluripotency and Human Primordial Germ Cell Development.

HAO, Y., CUI, Y. & GU, X. 2016. Genome-wide DNA methylation profiles changes associated with constant heat stress in pigs as measured by bisulfite sequencing. *Scientific reports,* 6.

HARGAN-CALVOPINA, J., TAYLOR, S., COOK, H., HU, Z., LEE, SERENA A., YEN, M.-R., CHIANG, Y.-S., CHEN, P.-Y. & CLARK, AMANDER T. 2016. Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Developmental Cell,* 39**,** 75-86.

HARR, J. C., LUPERCHIO, T. R., WONG, X., COHEN, E., WHEELAN, S. J. & REDDY, K. L. 2015. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *Journal of Cell Biology,* 208**,** 33-52.

HARVEY, A., CARETTI, G., MORESI, V., RENZINI, A. & ADAMO, S. 2019. Interplay between Metabolites and the Epigenome in Regulating Embryonic and Adult Stem Cell Potency and Maintenance. *Stem cell reports,* 13**,** 573-589.

HASSOUN, R., SCHWARTZ, P., RATH, D., VIEBAHN, C. & MÄNNER, J. 2010. Germ layer differentiation during early hindgut and cloaca formation in rabbit and pig embryos. *Journal of Anatomy,* 217**,** 665-678.

HAWKINS, R. D., HON, G. C., LEE, L. K., NGO, Q., LISTER, R., PELIZZOLA, M., EDSALL, L. E., KUAN, S., LUU, Y. & KLUGMAN, S. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell,* 6**,** 479-491.

HAYASHI, K., CHUVA DE SOUSA LOPES, S. M., KANEDA, M., TANG, F., HAJKOVA, P., LAO, K., O'CARROLL, D., DAS, P. P.,

TARAKHOVSKY, A., MISKA, E. A. & SURANI, M. A. 2008. MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis. *PLoS ONE,* 3**,** e1738.

HAYASHI, K., OGUSHI, S., KURIMOTO, K., SHIMAMOTO, S., OHTA, H. & 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. & SAITOU, M. 2011. Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. *Cell,* 146**,** 519-532.

HAYASHI, K. & SURANI, M. A. 2009. Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. *Development,* 136**,** 3549-3556.

HAYASHI, Y., OTSUKA, K., EBINA, M., IGARASHI, K., TAKEHARA, A., MATSUMOTO, M., KANAI, A., IGARASHI, K., SOGA, T. & MATSUI, Y. 2017. Distinct requirements for energy metabolism in mouse primordial germ cells and their reprogramming to embryonic germ cells. *Proceedings of the National Academy of Sciences,* 114**,** 8289-8294.

HE, Y.-F., LI, B.-Z., LI, Z., LIU, P., WANG, Y., TANG, Q., DING, J., JIA, Y., CHEN, Z., LI, L., SUN, Y., LI, X., DAI, Q., SONG, C.-X., ZHANG, K., HE, C. & XU, G.-L. 2011. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science (New York, N.Y.),* 333**,** 1303-1307.

HEARD, E. & MARTIENSSEN, R. A. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell,* 157**,** 95-109.

HEARD, E., ROUGEULLE, C., ARNAUD, D., AVNER, P., ALLIS, C. D. & SPECTOR, D. L. 2001. Methylation of Histone H3 at Lys-9 Is an Early Mark on the X Chromosome during X Inactivation. *Cell,* 107**,** 727-738.

HENCKEL, A., NAKABAYASHI, K., SANZ, L. A., FEIL, R., HATA, K. & ARNAUD, P. 2009. Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals. *Human Molecular Genetics,* 18**,** 3375-3383.

HENDRICH, B., GUY, J., RAMSAHOYE, B., WILSON, V. A. & BIRD, A. 2001. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes & Development,* 15**,** 710-723.

HERAS, S., SMITS, K., DE SCHAUWER, C. & VAN SOOM, A. 2017. Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during pronuclear development in equine zygotes produced by ICSI. *Epigenetics & Chromatin,* 10**,** 13.

HIKABE, O., HAMAZAKI, N., NAGAMATSU, G., OBATA, Y., HIRAO, Y., HAMADA, N., SHIMAMOTO, S., IMAMURA, T., NAKASHIMA, K., SAITOU, M. & HAYASHI, K. 2016. Reconstitution in vitro of the entire cycle of the mouse female germ line. *Nature,* 539**,** 299-303.

HILL, P. W. S., LEITCH, H. G., REQUENA, C. E., SUN, Z., AMOUROUX, R., ROMAN-TRUFERO, M., BORKOWSKA, M., TERRAGNI, J., VAISVILA, R., LINNETT, S., BAGCI, H., DHARMALINGHAM, G., HABERLE, V., LENHARD, B., ZHENG, Y., PRADHAN, S. & HAJKOVA, P. 2018. Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature,* 555**,** 392.

HILL, R. J. & CROSSAN, G. P. 2019. DNA cross-link repair safeguards genomic stability during premeiotic germ cell development. *Nature Genetics,* 51**,** 1283-1294.

HIRASAWA, R., CHIBA, H., KANEDA, M., TAJIMA, S., LI, E., JAENISCH, R. & SASAKI, H. 2008. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes & Development,* 22**,** 1607-1616.

HODGE, D. Q., CUI, J., GAMBLE, M. J. & GUO, W. 2018. Histone Variant MacroH2A1 Plays an Isoform-Specific Role in Suppressing Epithelial-Mesenchymal Transition. *Scientific Reports,* 8**,** 841.

HOEI-HANSEN, C. E., NIELSEN, J. E., ALMSTRUP, K., SONNE, S. B., GRAEM, N., SKAKKEBAEK, N. E., LEFFERS, H. & MEYTS, E. R.-D. 2004. Transcription Factor AP-2γ Is a Developmentally Regulated Marker of Testicular Carcinoma In situ and Germ Cell Tumors. *Clinical Cancer Research,* 10**,** 8521-8530.

HOLMES, E. E., JUNG, M., MELLER, S., LEISSE, A., SAILER, V., ZECH, J., MENGDEHL, M., GARBE, L.-A., UHL, B., KRISTIANSEN, G. & DIETRICH, D. 2014. Performance Evaluation of Kits for Bisulfite-Conversion of DNA from Tissues, Cell Lines, FFPE Tissues, Aspirates, Lavages, Effusions, Plasma, Serum, and Urine. *PLoS ONE,* 9**,** e93933.

HOLOCH, D. & MARGUERON, R. 2017. Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends in Biochemical Sciences*.

HON, GARY C., SONG, C.-X., DU, T., JIN, F., SELVARAJ, S., LEE, AH Y., YEN, C.-A., YE, Z., MAO, S.-Q., WANG, B.-A., KUAN, S., EDSALL, LEE E., ZHAO, BOXUAN S., XU, G.-L., HE, C. & REN, B. 2014. 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. *Molecular Cell,* 56**,** 286-297.

HOPF, C., VIEBAHN, C. & PÜSCHEL, B. 2011. BMP signals and the transcriptional repressor BLIMP1 during germline segregation in the mammalian embryo. *Development Genes and Evolution,* 221**,** 209.

HORSTHEMKE, B. 2018. A critical view on transgenerational epigenetic inheritance in humans. *Nature Communications,* 9**,** 2973.

HOYER-FENDER, S., COSTANZI, C. & PEHRSON, J. R. 2000. Histone MacroH2A1.2 Is Concentrated in the XY-Body by the Early Pachytene Stage of Spermatogenesis. *Experimental Cell Research,* 258**,** 254-260.

HSIAO, C. J., TUNG, P., BLISCHAK, J. D., BURNETT, J. E., BARR, K. A., DEY, K. K., STEPHENS, M. & GILAD, Y. 2020. Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. *Genome research,* 30**,** 611-621.

HU, S., LI, J., XU, F., MEI, S., LE DUFF, Y., YIN, L., PANG, X., CEN, S., JIN, Q. & LIANG, C. 2015. SAMHD1 inhibits LINE-1 retrotransposition by promoting stress granule formation. *PLoS genetics,* 11**,** e1005367.

HU, X., ZHANG, L., MAO, S.-Q., LI, Z., CHEN, J., ZHANG, R.-R., WU, H.-P., GAO, J., GUO, F. & LIU, W. 2014. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell stem cell,* 14**,** 512-522.

HUANG, J., ZHANG, H., WANG, X., DOBBS, K. B., YAO, J., QIN, G., WHITWORTH, K., WALTERS, E. M., PRATHER, R. S. & ZHAO, J.

2015. Impairment of preimplantation porcine embryo development by histone demethylase KDM5B knockdown through disturbance of bivalent H3K4me3-H3K27me3 modifications. *Biology of reproduction*, biolreprod. 114.122762.

HUANG, Y., CHAVEZ, L., CHANG, X., WANG, X., PASTOR, W. A., KANG, J., ZEPEDA-MARTÍNEZ, J. A., PAPE, U. J., JACOBSEN, S. E., PETERS, B. & RAO, A. 2014. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences,* 111**,** 1361-1366.

HURTADO-BAGÈS, S., GUBEROVIC, I. & BUSCHBECK, M. 2018. The MacroH2A1.1 – PARP1 Axis at the Intersection Between Stress Response and Metabolism. *Frontiers in Genetics,* 9.

HURTADO-BAGÈS, S., KNOBLOCH, G., LADURNER, A. G. & BUSCHBECK, M. 2020. The taming of PARP1 and its impact on NAD+ metabolism. *Molecular Metabolism,* 38**,** 100950.

HUYPENS, P., SASS, S., WU, M., DYCKHOFF, D., TSCHOP, M., THEIS, F., MARSCHALL, S., DE ANGELIS, M. H. & BECKERS, J. 2016. Epigenetic germline inheritance of diet-induced obesity and insulin resistance. *Nat Genet,* 48**,** 497-499.

HWANG, J. Y., OH, J.-N., PARK, C.-H., LEE, D.-K. & LEE, C.-K. 2015. Dosage compensation of X-chromosome inactivation center-linked genes in porcine preimplantation embryos: Non-chromosome-wide initiation of X-chromosome inactivation in blastocysts. *Mechanisms of Development,* 138**,** 246-255.

HYLDIG, S. M., CROXALL, N., CONTRERAS, D. A., THOMSEN, P. D. & ALBERIO, R. 2011a. Epigenetic reprogramming in the porcine germ line. *BMC Developmental Biology,* 11**,** 1-11.

HYLDIG, S. M. W., OSTRUP, O., VEJLSTED, M. & THOMSEN, P. D. 2011b. Changes of DNA Methylation Level and Spatial Arrangement of Primordial Germ Cells in Embryonic Day 15 to Embryonic Day 28 Pig Embryos. *Biology of Reproduction,* 84**,** 1087-1093.

ILICIC, T., KIM, J. K., KOLODZIEJCZYK, A. A., BAGGER, F. O., MCCARTHY, D. J., MARIONI, J. C. & TEICHMANN, S. A. 2016. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol,* 17**,** 29.

INOUE, A., JIANG, L., LU, F., SUZUKI, T. & ZHANG, Y. 2017. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature,* 547**,** 419-424.

INOUE, A. & ZHANG, Y. 2011. Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. *Science,* 334**,** 194-194.

IQBAL, K., TRAN, D. A., LI, A. X., WARDEN, C., BAI, A. Y., SINGH, P., WU, X., PFEIFER, G. P. & SZABÓ, P. E. 2015. Deleterious effects of endocrine disruptors are corrected in the mammalian germline by epigenome reprogramming. *Genome Biology,* 16**,** 59.

IRIE, N., WEINBERGER, L., TANG, WALFRED W., KOBAYASHI, T., VIUKOV, S., MANOR, Y. S., DIETMANN, S., HANNA, JACOB H. & SURANI, M A. 2015. SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. *Cell,* 160**,** 253-268.

ITO, S., D/'ALESSIO, A. C., TARANOVA, O. V., HONG, K., SOWERS, L. C. & 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.

IVANOVA, E., CANOVAS, S., GARCIA-MARTÍNEZ, S., ROMAR, R., LOPES, J. S., RIZOS, D., SANCHEZ-CALABUIG, M. J., KRUEGER, F., ANDREWS, S., PEREZ-SANZ, F., KELSEY, G. & COY, P. 2020. DNA methylation changes during preimplantation development reveal inter-species differences and reprogramming events at imprinted genes. *Clinical epigenetics,* 12**,** 64-64.

JAFARPOUR, F., HOSSEINI, S. M., OSTADHOSSEINI, S., ABBASI, H., DALMAN, A. & NASR-ESFAHANI, M. H. 2017. Comparative dynamics of 5-methylcytosine reprogramming and TET family expression during preimplantation mammalian development in mouse and sheep. *Theriogenology,* 89**,** 86-96.

JANG, H. & SERRA, C. 2014. Nutrition, Epigenetics, and Diseases. *Clinical Nutrition Research,* 3**,** 1-8.

JELTSCH, A. & JURKOWSKA, R. Z. 2014. New concepts in DNA methylation. *Trends in Biochemical Sciences,* 39**,** 310-318.

JIANG, H., LEI, R., DING, S. W. & ZHU, S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics,* 15**,** 182.

JIANG, S. & BALTIMORE, D. 2016. RNA-binding protein Lin28 in cancer and immunity. *Cancer Letters,* 375**,** 108-113.

JIH, G., IGLESIAS, N., CURRIE, M. A., BHANU, N. V., PAULO, J. A., GYGI, S. P., GARCIA, B. A. & MOAZED, D. 2017. Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. *Nature*.

JIMENEZ-CHILLARON, J. C., RAMON-KRAUEL, M., RIBO, S. & DIAZ, R. 2015. Transgenerational epigenetic inheritance of diabetes risk as a consequence of early nutritional imbalances. *Proceedings of the Nutrition Society,* 75**,** 78-89.

JONES, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.,* 13.

JONES, P. A. & TAKAI, D. 2001. The Role of DNA Methylation in Mammalian Epigenetics. *Science,* 293**,** 1068-1070.

JONKERS, I., MONKHORST, K., RENTMEESTER, E., GROOTEGOED, J. A., GROSVELD, F. & GRIBNAU, J. 2008. Xist RNA Is Confined to the Nuclear Territory of the Silenced X Chromosome throughout the Cell Cycle. *Molecular and Cellular Biology,* 28**,** 5583-5594.

JÖNSSON, M. E., LUDVIK BRATTÅS, P., GUSTAFSSON, C., PETRI, R., YUDOVICH, D., PIRCS, K., VERSCHUERE, S., MADSEN, S., HANSSON, J., LARSSON, J., MÅNSSON, R., MEISSNER, A. & JAKOBSSON, J. 2019. Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nature Communications,* 10**,** 3182.

JOSTES, S., FELLERMEYER, M., MERGES, G., KRISTIANSEN, G., NETTERSHEIM, D. & SCHORLE, H. 2019. Unique and redundant roles of SOX2 and SOX17 in regulating the germ cell tumour fate. *bioRxiv***,** 643114.

JOSTES, S. & SCHORLE, H. 2018. Signals and transcription factors for specification of human germ cells. *Stem cell investigation,* 5**,** 13-13.

JURKOWSKA, R. Z., JURKOWSKI, T. P. & JELTSCH, A. 2011a. Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem,* 12**,** 206-222.

JURKOWSKA, R. Z., RAJAVELU, A., ANSPACH, N., URBANKE, C., JANKEVICIUS, G., RAGOZIN, S., NELLEN, W. & JELTSCH, A. 2011b. Oligomerization and Binding of the Dnmt3a DNA Methyltransferase to Parallel DNA Molecules: HETEROCHROMATIC LOCALIZATION AND ROLE OF Dnmt3L. *The Journal of Biological Chemistry,* 286**,** 24200-24207.

KAFER, GEORGIA R., LI, X., HORII, T., SUETAKE, I., TAJIMA, S., HATADA, I. & CARLTON, PETER M. 2016. 5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability. *Cell Reports,* 14**,** 1283-1292.

KAGIWADA, S., KURIMOTO, K., HIROTA, T., YAMAJI, M. & SAITOU, M. 2013. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *The EMBO Journal,* 32**,** 340-353.

KANEHISA, M. & GOTO, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research,* 28**,** 27-30.

KANEHISA, M., SATO, Y., FURUMICHI, M., MORISHIMA, K. & TANABE, M. 2018. New approach for understanding genome variations in KEGG. *Nucleic Acids Research,* 47**,** D590-D595.

KARIMI, M. M., GOYAL, P., MAKSAKOVA, I. A., BILENKY, M., LEUNG, D., TANG, J. X., SHINKAI, Y., MAGER, D. L., JONES, S. & HIRST, M. 2011. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell stem cell,* 8**,** 676-687.

KASPAR, D., HASTREITER, S., IRMLER, M., HRABÉ DE ANGELIS, M. & BECKERS, J. 2020. Nutrition and its role in epigenetic inheritance of obesity and diabetes across generations. *Mammalian Genome***,** 1-15.

KATO, Y., ALAVATTAM, K. G., SIN, H.-S., MEETEI, A. R., PANG, Q., ANDREASSEN, P. R. & NAMEKAWA, S. H. 2015. FANCB is essential in the male germline and regulates H3K9 methylation on the sex chromosomes during meiosis. *Human molecular genetics,* 24**,** 5234-5249.

KAWASAKI, Y., LEE, J., MATSUZAWA, A., KOHDA, T., KANEKO-ISHINO, T. & ISHINO, F. 2014. Active DNA demethylation is required for complete imprint erasure in primordial germ cells. *Scientific Reports,* 4**,** 3658.

KAYA-OKUR, H. S., WU, S. J., CODOMO, C. A., PLEDGER, E. S., BRYSON, T. D., HENIKOFF, J. G., AHMAD, K. & HENIKOFF, S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications,* 10**,** 1930.

KEHLER, J., TOLKUNOVA, E., KOSCHORZ, B., PESCE, M., GENTILE, L., BOIANI, M., LOMELÍ, H., NAGY, A., MCLAUGHLIN, K. J. & SCHÖLER, H. R. 2004. Oct4 is required for primordial germ cell survival. *EMBO reports,* 5**,** 1078-1083.

KEMP, C. R., WILLEMS, E., WAWRZAK, D., HENDRICKX, M., AGBOR AGBOR, T. & LEYNS, L. 2007. Expression of Frizzled5, Frizzled7,

and Frizzled10 during early mouse development and interactions with canonical Wnt signaling. *Developmental Dynamics,* 236**,** 2011-2019.
KESTER, L. & VAN OUDENAARDEN, A. 2018. Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell,* 23**,** 166-179.
KEYS, J. L. & KING, G. J. 1989. Structural changes in the luminal epithelium of the porcine uterus between days 10 and 19 of the estrous cycle. *Am J Anat,* 185**,** 42-57.
KILLIAN, J. K., DORSSERS, L. C., TRABERT, B., GILLIS, A. J., COOK, M. B., WANG, Y., WATERFALL, J. J., STEVENSON, H., SMITH, W. I. & NOYES, N. 2016. Imprints and DPPA3 are bypassed during pluripotency-and differentiation-coupled methylation reprogramming in testicular germ cell tumors. *Genome Research,* 26**,** 1490-1504.
KIM, D.-H., TANG, Z., SHIMADA, M., FIERZ, B., HOUCK-LOOMIS, B., BAR-DAGEN, M., LEE, S., LEE, S.-K., MUIR, T. W., ROEDER, R. G. & LEE, J. W. 2013a. Histone H3K27 Trimethylation Inhibits H3 Binding and Function of SET1-Like H3K4 Methyltransferase Complexes. *Molecular and Cellular Biology,* 33**,** 4936-4946.
KIM, J., SHIN, Y., LEE, S., KIM, M., PUNJ, V., LU, J. F., SHIN, H., KIM, K., ULMER, T. S., KOH, J., JEONG, D. & AN, W. 2018a. Regulation of Breast Cancer-Induced Osteoclastogenesis by MacroH2A1.2 Involving EZH2-Mediated H3K27me3. *Cell Reports,* 24**,** 224-237.
KIM, J., STURGILL, D., SEBASTIAN, R., KHURANA, S., TRAN, A. D., EDWARDS, G. B., KRUSWICK, A., BURKETT, S., HOSOGANE, E. K., HANNON, W. W., WEYEMI, U., BONNER, W. M., LUGER, K. & OBERDOERFFER, P. 2018b. Replication Stress Shapes a Protective Chromatin Environment across Fragile Genomic Regions. *Molecular Cell,* 69**,** 36-47.e7.
KIM, J. M., HEO, K., CHOI, J., KIM, K. & AN, W. 2013b. The histone variant MacroH2A regulates Ca2+ influx through TRPC3 and TRPC6 channels. *Oncogenesis,* 2**,** e77-e77.
KIM, M., PARK, Y.-K., KANG, T.-W., LEE, S.-H., RHEE, Y.-H., PARK, J.-L., KIM, H.-J., LEE, D., LEE, D., KIM, S.-Y. & KIM, Y. S. 2014. Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. *Human Molecular Genetics,* 23**,** 657-667.
KIM, Y.-H., KIM, B.-J., KIM, B.-G., LEE, Y.-A., KIM, K.-J., CHUNG, H.-J., HWANG, S., WOO, J.-S., PARK, J.-K., SCHMIDT, J. A., PANG, M.-G. & RYU, B.-Y. 2013c. Stage-specific embryonic antigen-1 expression by undifferentiated spermatogonia in the prepubertal boar testis1. *Journal of Animal Science,* 91**,** 3143-3154.
KINKLEY, S., HELMUTH, J., POLANSKY, J. K., DUNKEL, I., GASPARONI, G., FRÖHLER, S., CHEN, W., WALTER, J., HAMANN, A. & CHUNG, H.-R. 2016. reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4+ memory T cells. *Nature Communications,* 7**,** 12514.
KLISCH, K., CONTRERAS, D. A., SUN, X., BREHM, R., BERGMANN, M. & ALBERIO, R. 2011. The Sda/GM2-glycan is a carbohydrate marker of porcine primordial germ cells and of a subpopulation of spermatogonia in cattle, pigs, horses and llama. *Reproduction,* 142**,** 667-74.

KOBAYASHI, H., SAKURAI, T., IMAI, M., TAKAHASHI, N., FUKUDA, A., YAYOI, O., SATO, S., NAKABAYASHI, K., HATA, K. & SOTOMARU, Y. 2012. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS genet,* 8**,** e1002440.

KOBAYASHI, H., SAKURAI, T., MIURA, F., IMAI, M., MOCHIDUKI, K., YANAGISAWA, E., SAKASHITA, A., WAKAI, T., SUZUKI, Y., ITO, T., MATSUI, Y. & KONO, T. 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research,* 23**,** 616-627.

KOBAYASHI, T. & SURANI, M. A. 2018. On the origin of the human germline. *Development,* 145**,** dev150433.

KOBAYASHI, T., ZHANG, H., TANG, W. W. C., IRIE, N., WITHEY, S., KLISCH, D., SYBIRNA, A., DIETMANN, S., CONTRERAS, D. A., WEBB, R., ALLEGRUCCI, C., ALBERIO, R. & SURANI, M. A. 2017. Principles of early human development and germ cell program from conserved model systems. *Nature,* advance online publication.

KOHLI, R. M. & ZHANG, Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature,* 502**,** 472.

KOJIMA, Y., SASAKI, K., YOKOBAYASHI, S., SAKAI, Y., NAKAMURA, T., YABUTA, Y., NAKAKI, F., NAGAOKA, S., WOLTJEN, K., HOTTA, A., YAMAMOTO, T. & SAITOU, M. 2017. Evolutionarily Distinctive Transcriptional and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. *Cell Stem Cell,* 21**,** 517-532.e5.

KONG, L., TAN, L., LV, R., SHI, Z., XIONG, L., WU, F., RABIDOU, K., SMITH, M., HE, C., ZHANG, L., QIAN, Y., MA, D., LAN, F., SHI, Y. & SHI, Y. G. 2016. A primary role of TET proteins in establishment and maintenance of De Novo bivalency at CpG islands. *Nucleic Acids Research,* 44**,** 8682-8692.

KOPPER, O. & BENVENISTY, N. 2012. Stepwise differentiation of human embryonic stem cells into early endoderm derivatives and their molecular characterization. *Stem Cell Research,* 8**,** 335-345.

KOZLOWSKI, M., CORUJO, D., HOTHORN, M., GUBEROVIC, I., MANDEMAKER, I. K., BLESSING, C., SPORN, J., GUTIERREZ‐TRIANA, A., SMITH, R., PORTMANN, T., TREIER, M., SCHEFFZEK, K., HUET, S., TIMINSZKY, G., BUSCHBECK, M. & LADURNER, A. G. 2018. MacroH2A histone variants limit chromatin plasticity through two distinct mechanisms. *EMBO reports,* 19.

KRASICH, R. & COPELAND, W. C. 2017. DNA polymerases in the mitochondria: A critical review of the evidence. *Frontiers in bioscience (Landmark edition),* 22**,** 692-709.

KRAUSS-ETSCHMANN, S., MEYER, K. F., DEHMEL, S. & HYLKEMA, M. N. 2015. Inter-and transgenerational epigenetic inheritance: evidence in asthma and COPD? *Clinical epigenetics,* 7**,** 53.

KREMSKY, I. & CORCES, V. G. 2020. Protection from DNA re-methylation by transcription factors in primordial germ cells and pre-implantation embryos can explain trans-generational epigenetic inheritance. *Genome Biology,* 21**,** 118.

KRENDL, C., SHAPOSHNIKOV, D., RISHKO, V., ORI, C., ZIEGENHAIN, C., SASS, S., SIMON, L., MÜLLER, N. S., STRAUB, T., BROOKS, K. E., CHAVEZ, S. L., ENARD, W., THEIS, F. J. & DRUKKER, M. 2017. GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. *Proceedings of the National Academy of Sciences,* 114**,** E9579-E9588.

KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics,* 27**,** 1571-2.

KUMAR, A. S., SEAH, M. K. Y., LING, K. Y., WANG, Y., TAN, J. H. L., NITSCH, S., LIM, S. L., LORTHONGPANICH, C., WOLLMANN, H., LOW, D. H. P., GUCCIONE, E. & MESSERSCHMIDT, D. M. 2017. Loss of maternal Trim28 causes male-predominant early embryonic lethality. *Genes & Development*.

KUMAR, R. P., RAY, S., HOME, P., SAHA, B., BHATTACHARYA, B., WILKINS, H. M., CHAVAN, H., GANGULY, A., MILANO-FOSTER, J. & PAUL, A. 2018. Regulation of energy metabolism during early mammalian development: TEAD4 controls mitochondrial transcription. *Development,* 145**,** dev162644.

KURAMOCHI-MIYAGAWA, S., WATANABE, T., GOTOH, K., TAKAMATSU, K., CHUMA, S., KOJIMA-KITA, K., SHIROMOTO, Y., ASADA, N., TOYODA, A. & FUJIYAMA, A. 2010. MVH in piRNA processing and gene silencing of retrotransposons. *Genes & development,* 24**,** 887-892.

KURIMOTO, K., YABUTA, Y., HAYASHI, K., OHTA, H., KIYONARI, H., MITANI, T., MORITOKI, Y., KOHRI, K., KIMURA, H., YAMAMOTO, T., KATOU, Y., SHIRAHIGE, K. & SAITOU, M. 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. & SAITOU, M. 2008. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes & development,* 22**,** 1617-1635.

KUSTATSCHER, G., HOTHORN, M., PUGIEUX, C., SCHEFFZEK, K. & LADURNER, A. G. 2005. Splicing regulates NAD metabolite binding to histone macroH2A. *Nature Structural & Molecular Biology,* 12**,** 624-625.

LA MANNO, G., SOLDATOV, R., ZEISEL, A., BRAUN, E., HOCHGERNER, H., PETUKHOV, V., LIDSCHREIBER, K., KASTRITI, M. E., LÖNNERBERG, P., FURLAN, A., FAN, J., BORM, L. E., LIU, Z., VAN BRUGGEN, D., GUO, J., HE, X., BARKER, R., SUNDSTRÖM, E., CASTELO-BRANCO, G., CRAMER, P., ADAMEYKO, I., LINNARSSON, S. & KHARCHENKO, P. V. 2018. RNA velocity of single cells. *Nature,* 560**,** 494-498.

LADURNER, A. G. 2003. Inactivating Chromosomes: A macro Domain that Minimizes Transcription. *Molecular Cell,* 12**,** 1-3.

LAM, A. Q., FREEDMAN, B. S., MORIZANE, R., LEROU, P. H., VALERIUS, M. T. & BONVENTRE, J. V. 2014. Rapid and efficient differentiation of human pluripotent stem cells into intermediate mesoderm that forms tubules expressing kidney proximal tubular markers. *Journal of the American Society of Nephrology : JASN,* 25**,** 1211-1225.

LARRIBA, E. & DEL MAZO, J. 2016. Role of Non-Coding RNAs in the Transgenerational Epigenetic Transmission of the Effects of Reprotoxicants. *International Journal of Molecular Sciences,* 17**,** 452.

LAWSON, K. & HAGE, W. 1994. Clonal analysis of the origin of primordial germ cells in the mouse. *Germline development,* 165**,** 68-84.

LE, Q., YAN, B., YU, X., LI, Y., SONG, H., ZHU, H., HOU, W., MA, D., WU, F., ZHOU, Y. & MA, L. 2017. Drug-seeking motivation level in male rats determines offspring susceptibility or resistance to cocaine-seeking behaviour. 8**,** 15527.

LEE CHONG, T., AHEARN, E. L. & CIMMINO, L. 2019. Reprogramming the Epigenome With Vitamin C. *Frontiers in Cell and Developmental Biology,* 7.

LEE, K., HAMM, J., WHITWORTH, K., SPATE, L., PARK, K.-W., MURPHY, C. N. & PRATHER, R. S. 2014a. Dynamics of TET family expression in porcine preimplantation embryos is related to zygotic genome activation and required for the maintenance of NANOG. *Developmental Biology,* 386**,** 86-95.

LEE, M. T., BONNEAU, A. R. & GIRALDEZ, A. J. 2014b. Zygotic genome activation during the maternal-to-zygotic transition. *Annual review of cell and developmental biology,* 30**,** 581-613.

LEE, S.-T., MUENCH, M. O., FOMIN, M., XIAO, J., ZHOU, M., DE SMITH, A. J., MARTÍN-SUBERO, J., HEATH, S., HOUSEMAN, E. A., ROY, R., WRENSCH, M. R., WIENCKE, J. K., METAYER, C. & WIEMELS, J. L. 2014c. Two-Track Epigenetic Remodeling and Backtracking to Embryonic Stem Cell Bivalency in B-Cell Acute Lymphoblastic Leukemias. *Blood,* 124**,** 3557-3557.

LEGAULT, L.-M., BERTRAND-LEHOUILLIER, V. & MCGRAW, S. 2017. Preimplantation alcohol exposure and developmental programming of FASD: An epigenetic perspective. *Biochemistry and Cell Biology*.

LEOPARDO, N. P., INSERRA, P. I. & VITULLO, A. D. 2018. Challenging the Paradigms on the Origin, Specification and Development of the Female Germ Line in Placental Mammals. *Germ Cell***,** 101.

LEOPARDO, N. P. & VITULLO, A. D. 2017. Early embryonic development and spatiotemporal localization of mammalian primordial germ cell-associated proteins in the basal rodent Lagostomus maximus. *Scientific Reports,* 7**,** 594.

LESCH, B. J., DOKSHIN, G. A., YOUNG, R. A., MCCARREY, J. R. & PAGE, D. C. 2013. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 16061-16066.

LESCH, B. J., SILBER, S. J., MCCARREY, J. R. & PAGE, D. C. 2016. Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat Genet,* 48**,** 888-894.

LEVINE, A. J. & BRIVANLOU, A. H. 2006. GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. *Development,* 133**,** 209-216.

LI, H., LIEFKE, R., JIANG, J., KURLAND, J. V., TIAN, W., DENG, P., ZHANG, W., HE, Q., PATEL, D. J., BULYK, M. L., SHI, Y. & WANG, Z. 2017a. Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature,* 549**,** 287-291.

LI, L., DONG, J., YAN, L., YONG, J., LIU, X., HU, Y., FAN, X., WU, X., GUO, H. & WANG, X. 2017b. Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell*.

LI, L., LIU, C., BIECHELE, S., ZHU, Q., SONG, L., LANNER, F., JING, N. & ROSSANT, J. 2013a. Location of transient ectodermal progenitor potential in mouse development. *Development,* 140**,** 4533-4543.

LI, Y., MORETTO-ZITA, M., SONCIN, F., WAKELAND, A., WOLFE, L., LEON-GARCIA, S., PANDIAN, R., PIZZO, D., CUI, L., NAZOR, K., LORING, J. F., CRUM, C. P., LAURENT, L. C. & PARAST, M. M. 2013b. BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a ΔNp63⁺ cytotrophoblast stem cell state. *Development,* 140**,** 3965-3976.

LI, Z., FANG, F., ZHAO, Q., LI, H. & XIONG, C. 2019. Supplementation of vitamin C promotes early germ cell specification from human embryonic stem cells. *Stem cell research & therapy,* 10**,** 324-324.

LI, Z., GU, T.-P., WEBER, A. R., SHEN, J.-Z., LI, B.-Z., XIE, Z.-G., YIN, R., GUO, F., LIU, X. & TANG, F. 2015. Gadd45a promotes DNA demethylation through TDG. *Nucleic acids research,* 43**,** 3986-3997.

LIAO, Y., SMYTH, G. K. & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

LIM, J. P. & BRUNET, A. 2013. Bridging the transgenerational gap with epigenetic memory. *Trends in genetics : TIG,* 29**,** 176-186.

LIM, S. L., QU, Z. P., KORTSCHAK, R. D., LAWRENCE, D. M., GEOGHEGAN, J., HEMPFLING, A.-L., BERGMANN, M., GOODNOW, C. C., ORMANDY, C. J. & WONG, L. 2015. HENMT1 and piRNA stability are required for adult male germ cell transposon repression and to define the spermatogenic program in the mouse. *PLoS Genet,* 11**,** e1005620.

LIN, I. H., CHEN, Y.-F. & HSU, M.-T. 2017. Correlated 5-Hydroxymethylcytosine (5hmC) and Gene Expression Profiles Underpin Gene and Organ-Specific Epigenetic Regulation in Adult Mouse Brain and Liver. *PLOS ONE,* 12**,** e0170779.

LIN, Y. & PAGE, D. C. 2005. Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. *Developmental Biology,* 288**,** 309-316.

LINDROTH, A. M., PARK, Y. J., MCLEAN, C. M., DOKSHIN, G. A., PERSSON, J. M., HERMAN, H., PASINI, D., MIRO, X., DONOHOE, M. E. & LEE, J. T. 2008. Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. *PLoS Genet,* 4**,** e1000145.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z. & NGO, Q.-M.

2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *nature,* 462**,** 315-322.

LIU, C., WANG, R., HE, Z., OSTEIL, P., WILKIE, E., YANG, X., CHEN, J., CUI, G., GUO, W. & CHEN, Y. 2018. Suppressing nodal signaling activity predisposes ectodermal differentiation of epiblast stem cells. *Stem cell reports,* 11**,** 43-57.

LIU, S., BRIND'AMOUR, J., KARIMI, M. M., SHIRANE, K., BOGUTZ, A., LEFEBVRE, L., SASAKI, H., SHINKAI, Y. & LORINCZ, M. C. 2014. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & Development,* 28**,** 2041-2055.

LIU, X., GAO, Q., LI, P., ZHAO, Q., ZHANG, J., LI, J., KOSEKI, H. & WONG, J. 2013. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nature Communications,* 4**,** 1563.

LIU, X., WANG, C., LIU, W., LI, J., LI, C., KOU, X., CHEN, J., ZHAO, Y., GAO, H., WANG, H., ZHANG, Y., GAO, Y. & GAO, S. 2016. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature,* 537**,** 558-562.

LU, J. Y., SHAO, W., CHANG, L., YIN, Y., LI, T., ZHANG, H., HONG, Y., PERCHARDE, M., GUO, L., WU, Z., LIU, L., LIU, W., YAN, P., RAMALHO-SANTOS, M., SUN, Y. & SHEN, X. 2020. Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Reports,* 30**,** 3296-3311.e5.

LUO, Y., HARTFORD, S. A., ZENG, R., SOUTHARD, T. L., SHIMA, N. & SCHIMENTI, J. C. 2014. Hypersensitivity of primordial germ cells to compromised replication-associated DNA repair involves ATM-p53-p21 signaling. *PLoS Genet,* 10**,** e1004471.

LYNCH, M. D., SMITH, A. J., DE GOBBI, M., FLENLEY, M., HUGHES, J. R., VERNIMMEN, D., AYYUB, H., SHARPE, J. A., SLOANE‐STANLEY, J. A. & SUTHERLAND, L. 2012. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO journal,* 31**,** 317-329.

LYON, M. F. 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). *nature,* 190**,** 372.

MA, H., ZHAI, J., WAN, H., JIANG, X., WANG, X., WANG, L., XIANG, Y., HE, X., ZHAO, Z.-A., ZHAO, B., ZHENG, P., LI, L. & WANG, H. 2019. In vitro culture of cynomolgus monkey embryos beyond early gastrulation. *Science,* 366**,** eaax7890.

MA, L., BUCHOLD, G. M., GREENBAUM, M. P., ROY, A., BURNS, K. H., ZHU, H., HAN, D. Y., HARRIS, R. A., COARFA, C. & GUNARATNE, P. H. 2009. GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet,* 5**,** e1000635.

MAATOUK, D. M., KELLAM, L. D., MANN, M. R. W., LEI, H., LI, E., BARTOLOMEI, M. S. & RESNICK, J. L. 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.

MADISON, B. B. 2016. Srebp2: A master regulator of sterol and fatty acid synthesis. *Journal of lipid research,* 57**,** 333-335.

MAGNÚSDÓTTIR, E., DIETMANN, S., MURAKAMI, K., GÜNESDOGAN, U., TANG, F., BAO, S., DIAMANTI, E., LAO, K., GOTTGENS, B. & SURANI, M. A. 2013. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nature cell biology,* 15**,** 905-915.

MAITI, A. & DROHAT, A. C. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry,* 286**,** 35334-35338.

MAK, W., BAXTER, J., SILVA, J., NEWALL, A. E., OTTE, A. P. & BROCKDORFF, N. 2002. Mitotically Stable Association of Polycomb Group Proteins Eed and Enx1 with the Inactive X Chromosome in Trophoblast Stem Cells. *Current Biology,* 12**,** 1016-1020.

MAK, W., NESTEROVA, T. B., DE NAPOLES, M., APPANAH, R., YAMANAKA, S., OTTE, A. P. & BROCKDORFF, N. 2004. Reactivation of the Paternal X Chromosome in Early Mouse Embryos. *Science,* 303**,** 666-669.

MALLOL, A., GUIROLA, M. & PAYER, B. 2019. PRDM14 controls X-chromosomal and global epigenetic reprogramming of H3K27me3 in migrating mouse primordial germ cells. *Epigenetics & Chromatin,* 12**,** 38.

MANIKKAM, M., GUERRERO-BOSAGNA, C., TRACEY, R., HAQUE, M. M. & SKINNER, M. K. 2012. Transgenerational Actions of Environmental Compounds on Reproductive Disease and Identification of Epigenetic Biomarkers of Ancestral Exposures. *PLOS ONE,* 7**,** e31901.

MANKU, G. & CULTY, M. 2015. Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. *Reproduction,* 149**,** R139-R157.

MANSOUR, A. A., GAFNI, O., WEINBERGER, L., ZVIRAN, A., AYYASH, M., RAIS, Y., KRUPALNIK, V., ZERBIB, M., AMANN-ZALCENSTEIN, D., MAZA, I., GEULA, S., VIUKOV, S., HOLTZMAN, L., PRIBLUDA, A., CANAANI, E., HORN-SABAN, S., AMIT, I., NOVERSHTERN, N. & HANNA, J. H. 2012. The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature,* 488**,** 409-413.

MASALA, L., BURRAI, G. P., BELLU, E., ARIU, F., BOGLIOLO, L., LEDDA, S. & BEBBERE, D. 2017. Methylation dynamics during folliculogenesis and early embryo development in sheep. *Reproduction*.

MCKENNA, A., HANNA, M., BANKS, E., SIVACHENKO, A., CIBULSKIS, K., KERNYTSKY, A., GARIMELLA, K., ALTSHULER, D., GABRIEL, S. & DALY, M. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research,* 20**,** 1297-1303.

MCLAREN, A. & LAWSON, K. A. 2005. How is the mouse germ-cell lineage established? *Differentiation,* 73**,** 435-437.

MCMURCHY, A. N., STEMPOR, P., GAARENSTROOM, T., WYSOLMERSKI, B., DONG, Y., AUSSIANIKAVA, D., APPERT,

A., HUANG, N., KOLASINSKA-ZWIERZ, P., SAPETSCHNIG, A., MISKA, E. A. & AHRINGER, J. 2017. A team of heterochromatin factors collaborates with small RNA pathways to combat repetitive elements and germline stress. *eLife,* 6**,** e21666.

MEDEIROS, L. A., DENNIS, L. M., GILL, M. E., HOUBAVIY, H., MARKOULAKI, S., FU, D., WHITE, A. C., KIRAK, O., SHARP, P. A., PAGE, D. C. & JAENISCH, R. 2011. Mir-290–295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proceedings of the National Academy of Sciences,* 108**,** 14163-14168.

MEHROTRA, P. V., AHEL, D., RYAN, D. P., WESTON, R., WIECHENS, N., KRAEHENBUEHL, R., OWEN-HUGHES, T. & AHEL, I. 2011. DNA Repair Factor APLF Is a Histone Chaperone. *Molecular Cell,* 41**,** 46-55.

MEISSNER, A., MIKKELSEN, T. S., GU, H., WERNIG, M., HANNA, J., SIVACHENKO, A., ZHANG, X., BERNSTEIN, B. E., NUSBAUM, C. & JAFFE, D. B. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature,* 454**,** 766-770.

MELLEN, M., AYATA, P., DEWELL, S., KRIAUCIONIS, S. & HEINTZ, N. 2012. MeCP2 binds to 5hmc enriched within active genes and accessible chromatin in the nervous system. *Cell,* 151**,** 1417-1430.

MENDENHALL, E. M., KOCHE, R. P., TRUONG, T., ZHOU, V. W., ISSAC, B., CHI, A. S., KU, M. & BERNSTEIN, B. E. 2010. GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. *PLOS Genetics,* 6**,** e1001244.

MESSERSCHMIDT, D. M. 2012. Should I stay or should I go: Protection and maintenance of DNA methylation at imprinted genes. *Epigenetics,* 7**,** 969-975.

MINAJIGI, A., FROBERG, J. E., WEI, C., SUNWOO, H., KESNER, B., COLOGNORI, D., LESSING, D., PAYER, B., BOUKHALI, M., HAAS, W. & LEE, J. T. 2015. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science,* 349.

MIRA-BONTENBAL, H. & GRIBNAU, J. 2016. New Xist-Interacting Proteins in X-Chromosome Inactivation. *Current Biology,* 26**,** R338-R342.

MIRANDA-GONÇALVES, V., LAMEIRINHAS, A., HENRIQUE, R. & JERÓNIMO, C. 2018. Metabolism and Epigenetic Interplay in Cancer: Regulation and Putative Therapeutic Targets. *Frontiers in Genetics,* 9.

MITSUNAGA, S., ODAJIMA, J., YAWATA, S., SHIODA, K., OWA, C., ISSELBACHER, K. J., HANNA, J. H. & SHIODA, T. 2017. Relevance of iPSC-derived human PGC-like cells at the surface of embryoid bodies to prechemotaxis migrating PGCs. *Proceedings of the National Academy of Sciences,* 114**,** E9913-E9922.

MOCHIZUKI, K., HAYASHI, Y., SEKINAKA, T., OTSUKA, K., ITO-MATSUOKA, Y., KOBAYASHI, H., OKI, S., TAKEHARA, A., KONO, T., OSUMI, N. & MATSUI, Y. 2018a. Repression of Somatic Genes by Selective Recruitment of HDAC3 by BLIMP1 Is Essential for Mouse Primordial Germ Cell Fate Determination. *Cell Reports,* 24**,** 2682-2693.e6.

MOCHIZUKI, K., TANDO, Y., SEKINAKA, T., OTSUKA, K., HAYASHI, Y., KOBAYASHI, H., KAMIO, A., ITO-MATSUOKA, Y., TAKEHARA, A., KONO, T., OSUMI, N. & MATSUI, Y. 2018b. SETDB1 is essential for mouse primordial germ cell fate determination by ensuring BMP signaling. *Development,* 145**,** dev164160.

MOHN, F., WEBER, M., REBHAN, M., ROLOFF, T. C., RICHTER, J., STADLER, M. B., BIBEL, M. & SCHÜBELER, D. 2008. Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. *Molecular Cell,* 30**,** 755-766.

MOHNI, K. N., WESSEL, S. R., ZHAO, R., WOJCIECHOWSKI, A. C., LUZWICK, J. W., LAYDEN, H., EICHMAN, B. F., THOMPSON, P. S., MEHTA, K. P. M. & CORTEZ, D. 2019. HMCES Maintains Genome Integrity by Shielding Abasic Sites in Single-Strand DNA. *Cell,* 176**,** 144-153.e13.

MOLARO, A. & MALIK, H. S. 2016. Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline. *Current Opinion in Genetics & Development,* 37**,** 51-58.

MOLYNEAUX, K. A., STALLOCK, J., SCHAIBLE, K. & WYLIE, C. 2001. Time-Lapse Analysis of Living Mouse Germ Cell Migration. *Developmental Biology,* 240**,** 488-498.

MOLYNEAUX, K. A., ZINSZNER, H., KUNWAR, P. S., SCHAIBLE, K., STEBLER, J., SUNSHINE, M. J., O'BRIEN, W., RAZ, E., LITTMAN, D., WYLIE, C. & LEHMANN, R. 2003. The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. *Development,* 130**,** 4279-4286.

MONK, M. & HARPER, M. I. 1979. Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature,* 281**,** 311.

MOREIRA DE MELLO, J. C., FERNANDES, G. R., VIBRANOVSKI, M. D. & PEREIRA, L. V. 2017. Early X chromosome inactivation during human preimplantation development revealed by single-cell RNA-sequencing. *Scientific Reports,* 7**,** 10794.

MORISON, I. M., RAMSAY, J. P. & SPENCER, H. G. 2005. A census of mammalian imprinting. *Trends Genet,* 21**,** 457-65.

MULAS, C., KALKAN, T. & SMITH, A. 2017. NODAL Secures Pluripotency upon Embryonic Stem Cell Progression from the Ground State. *Stem Cell Reports,* 9**,** 77-91.

MUOTRI, A. R., CHU, V. T., MARCHETTO, M. C. N., DENG, W., MORAN, J. V. & GAGE, F. H. 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature,* 435**,** 903-910.

MURAKAMI, K., GÜNESDOGAN, U., ZYLICZ, J. J., TANG, W. W. C., SENGUPTA, R., KOBAYASHI, T., KIM, S., BUTLER, R., DIETMANN, S. & SURANI, M. A. 2016. NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature,* 529**,** 403-407.

MURATA, M. M., KONG, X., MONCADA, E., CHEN, Y., IMAMURA, H., WANG, P., BERNS, M. W., YOKOMORI, K. & DIGMAN, M. A. 2019. NAD+ consumption by PARP1 in response to DNA damage

triggers metabolic shift critical for damaged cell survival. *Molecular biology of the cell,* 30**,** 2584-2597.

MURDOCH, B. M., MURDOCH, G. K., GREENWOOD, S. & MCKAY, S. 2016. Nutritional Influence on Epigenetic Marks and Effect on Livestock Production. *Frontiers in Genetics,* 7**,** 182.

MUTZEL, V., OKAMOTO, I., DUNKEL, I., SAITOU, M., GIORGETTI, L., HEARD, E. & SCHULZ, E. G. 2019. A symmetric toggle switch explains the onset of random X inactivation in different mammals. *Nature Structural & Molecular Biology,* 26**,** 350-360.

NADLER, J. J. & BRAUN, R. E. 2000. Fanconi anemia complementation group C is required for proliferation of murine primordial germ cells.

NAIR, M., TENG, A., BILANCHONE, V., AGRAWAL, A., LI, B. & DAI, X. 2006. Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription. *The Journal of cell biology,* 173**,** 253-264.

NAKAKI, F., HAYASHI, K., OHTA, H., KURIMOTO, K., YABUTA, Y. & SAITOU, M. 2013. Induction of mouse germ-cell fate by transcription factors in vitro. *Nature,* 501**,** 222.

NAKAMURA, T., ARAI, Y., UMEHARA, H., MASUHARA, M., KIMURA, T., TANIGUCHI, H., SEKIMOTO, T., IKAWA, M., YONEDA, Y., OKABE, M., TANAKA, S., SHIOTA, K. & NAKANO, T. 2007. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat Cell Biol,* 9**,** 64-71.

NAKAMURA, T., LIU, Y.-J., NAKASHIMA, H., UMEHARA, H., INOUE, K., MATOBA, S., TACHIBANA, M., OGURA, A., SHINKAI, Y. & NAKANO, T. 2012. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature,* 486**,** 415-419.

NEKRUTENKO, A. & LI, W.-H. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics,* 17**,** 619-621.

NELLÅKER, C., KEANE, T. M., YALCIN, B., WONG, K., AGAM, A., BELGARD, T. G., FLINT, J., ADAMS, D. J., FRANKEL, W. N. & PONTING, C. P. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome biology,* 13**,** R45-R45.

NERI, F., INCARNATO, D., KREPELOVA, A., RAPELLI, S., PAGNANI, A., ZECCHINA, R., PARLATO, C. & OLIVIERO, S. 2013a. Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biology,* 14**,** R91.

NERI, F., KREPELOVA, A., INCARNATO, D., MALDOTTI, M., PARLATO, C., GALVAGNI, F., MATARESE, F., STUNNENBERG, HENDRIK G. & OLIVIERO, S. 2013b. Dnmt3L Antagonizes DNA Methylation at Bivalent Promoters and Favors DNA Methylation at Gene Bodies in ESCs. *Cell,* 155**,** 121-134.

NESTEROVA, T. B., MERMOUD, J. E., HILTON, K., PEHRSON, J., SURANI, M. A., MCLAREN, A. & BROCKDORFF, N. 2002. Xist expression and macroH2A1.2 localisation in mouse primordial and pluripotent embryonic germ cells. *Differentiation,* 69**,** 216-225.

NESTOR, C. E., OTTAVIANO, R., REDDINGTON, J., SPROUL, D., REINHARDT, D., DUNICAN, D., KATZ, E., DIXON, J. M., HARRISON, D. J. & MEEHAN, R. R. 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Research,* 22**,** 467-477.

NEWKIRK, S. J., LEE, S., GRANDI, F. C., GAYSINSKAYA, V., ROSSER, J. M., VANDEN BERG, N., HOGARTH, C. A., MARCHETTO, M. C. N., MUOTRI, A. R., GRISWOLD, M. D., YE, P., BORTVIN, A., GAGE, F. H., BOEKE, J. D. & AN, W. 2017. Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proceedings of the National Academy of Sciences,* 114**,** E5635-E5644.

NG, J.-H., KUMAR, V., MURATANI, M., KRAUS, P., YEO, J.-C., YAW, L.-P., XUE, K., LUFKIN, T., PRABHAKAR, S. & NG, H.-H. 2013. In Vivo Epigenomic Profiling of Germ Cells Reveals Germ Cell Molecular Signatures. *Developmental Cell,* 24**,** 324-333.

NIKOLIC, A., VOLAREVIC, V., ARMSTRONG, L., LAKO, M. & STOJKOVIC, M. Primordial Germ Cells: Current Knowledge and Perspectives. *Stem Cells International*.

NIU, Y., SUN, N., LI, C., LEI, Y., HUANG, Z., WU, J., SI, C., DAI, X., LIU, C., WEI, J., LIU, L., FENG, S., KANG, Y., SI, W., WANG, H., ZHANG, E., ZHAO, L., LI, Z., LUO, X., CUI, G., PENG, G., IZPISÚA BELMONTE, J. C., JI, W. & TAN, T. 2019. Dissecting primate early post-implantation development using long-term in vitro embryo culture. *Science,* 366**,** eaaw5754.

OESTRUP, O., HALL, V., PETKOV, S., WOLF, X., HYLDIG, S. & HYTTEL, P. 2009. From Zygote to Implantation: Morphological and Molecular Dynamics during Embryo Development in the Pig. *Reproduction in Domestic Animals,* 44**,** 39-49.

OHINATA, Y., OHTA, H., SHIGETA, M., YAMANAKA, K., WAKAYAMA, T. & 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. C., OBUKHANYCH, T., NUSSENZWEIG, M. & TARAKHOVSKY, A. 2005. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature,* 436**,** 207-213.

OHNO, R., NAKAYAMA, M., NARUSE, C., OKASHITA, N., TAKANO, O., TACHIBANA, M., ASANO, M., SAITOU, M. & SEKI, Y. 2013. A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development,* 140**,** 2892-2903.

OHNO, S. 2013. *Sex chromosomes and sex-linked genes*, Springer Science & Business Media.

OKAE, H., CHIBA, H., HIURA, H., HAMADA, H., SATO, A., UTSUNOMIYA, T., KIKUCHI, H., YOSHIDA, H., TANAKA, A., SUYAMA, M. & ARIMA, T. 2014. Genome-Wide Analysis of DNA Methylation Dynamics during Early Human Development. *PLoS Genetics,* 10**,** e1004868.

OKAMOTO, I., ARNAUD, D., LE BACCON, P., OTTE, A. P., DISTECHE, C. M., AVNER, P. & HEARD, E. 2005. Evidence for de novo

imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature,* 438**,** 369.

OKAMOTO, I., OTTE, A. P., ALLIS, C. D., REINBERG, D. & HEARD, E. 2004. Epigenetic Dynamics of Imprinted X Inactivation During Early Mouse Development. *Science,* 303**,** 644-649.

OKAMOTO, I., PATRAT, C., THÉPOT, D., PEYNOT, N., FAUQUE, P., DANIEL, N., DIABANGOUAYA, P., WOLF, J.-P., RENARD, J.-P., DURANTHON, V. & HEARD, E. 2011. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature,* 472**,** 370-374.

OOI, S. K. T., QIU, C., BERNSTEIN, E., LI, K., JIA, D., YANG, Z., ERDJUMENT-BROMAGE, H., TEMPST, P., LIN, S.-P., ALLIS, C. D., CHENG, X. & BESTOR, T. H. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature,* 448**,** 714-717.

OZAWA, M., FUKUDA, T., SAKAMOTO, R., HONDA, H. & YOSHIDA, N. 2016. The Histone Demethylase FBXL10 Regulates the Proliferation of Spermatogonia and Ensures Long-Term Sustainable Spermatogenesis in Mice1. *Biology of Reproduction,* 94**,** 92, 1-11-92, 1-11.

PAN, X., CANG, X., DAN, S., LI, J., CHENG, J., KANG, B., DUAN, X., SHEN, B. & WANG, Y.-J. 2016. Site-specific Disruption of the Oct4/Sox2 Protein Interaction Reveals Coordinated Mesendodermal Differentiation and the Epithelial-Mesenchymal Transition. *The Journal of biological chemistry,* 291**,** 18353-18369.

PANTIER, R., MULLIN, N., HALL-PONSELE, E. & CHAMBERS, I. 2020. TET1 interacts directly with NANOG via independent regions containing hydrophobic and aromatic residues. *bioRxiv***,** 2020.06.29.177584.

PARK, K.-E., JOHNSON, C. M., WANG, X. & CABOT, R. A. 2011. Differential developmental requirements for individual histone H3K9 methyltransferases in cleavage-stage porcine embryos. *Reproduction, Fertility and Development,* 23**,** 551-560.

PASQUE, V., GILLICH, A., GARRETT, N. & GURDON, J. B. 2011. Histone variant macroH2A confers resistance to nuclear reprogramming. *The EMBO Journal,* 30**,** 2373-2387.

PASQUE, V. & PLATH, K. 2015. X chromosome reactivation in reprogramming and in development. *Current Opinion in Cell Biology,* 37**,** 75-83.

PASQUE, V., RADZISHEUSKAYA, A., GILLICH, A., HALLEY-STOTT, R. P., PANAMAROVA, M., ZERNICKA-GOETZ, M., SURANI, M. A. & SILVA, J. C. R. 2012. Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. *Journal of cell science,* 125**,** 6094-6104.

PASQUE, V., TCHIEU, J., KARNIK, R., UYEDA, M., SADHU DIMASHKIE, A., CASE, D., PAPP, B., BONORA, G., PATEL, S., HO, R., SCHMIDT, R., MCKEE, R., SADO, T., TADA, T., MEISSNER, A. & PLATH, K. 2014. X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency. *Cell,* 159**,** 1681-1697.

PASTOR, W. A., PAPE, U. J., HUANG, Y., HENDERSON, H. R., LISTER, R., KO, M., MCLOUGHLIN, E. M., BRUDNO, Y., MAHAPATRA, S., KAPRANOV, P., TAHILIANI, M., DALEY, G. Q., LIU, X. S., ECKER, J. R., MILOS, P. M., AGARWAL, S. & RAO, A. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature,* 473**,** 394-397.

PASTOR, W. A., STROUD, H., NEE, K., LIU, W., PEZIC, D., MANAKOV, S., LEE, S. A., MOISSIARD, G., ZAMUDIO, N. & BOURC'HIS, D. 2014. MORC1 represses transposable elements in the mouse male germline. *Nature communications,* 5**,** 1-14.

PATEL, M. S. & KOROTCHKINA, L. G. 2001. Regulation of mammalian pyruvate dehydrogenase complex by phosphorylation: complexity of multiple phosphorylation sites and kinases. *Experimental & Molecular Medicine,* 33**,** 191-197.

PATEL, M. S. & ROCHE, T. E. 1990. Molecular biology and biochemistry of pyruvate dehydrogenase complexes. *The FASEB Journal,* 4**,** 3224-3233.

PATEL, S., BONORA, G., SAHAKYAN, A., KIM, R., CHRONIS, C., LANGERMAN, J., FITZ-GIBBON, S., RUBBI, L., SKELTON, R. J. P., ARDEHALI, R., PELLEGRINI, M., LOWRY, W. E., CLARK, A. T. & PLATH, K. 2017. Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. *Cell Reports,* 18**,** 54-67.

PATIL, D. P., CHEN, C.-K., PICKERING, B. F., CHOW, A., JACKSON, C., GUTTMAN, M. & JAFFREY, S. R. 2016. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature,* 537**,** 369-373.

PATTEN, B. M. 1931. The embryology of the pig / by Bradley M. Patten, Philadelphia, Philadelphia : Blakiston Company.

PAYER, B., ROSENBERG, M., YAMAJI, M., YABUTA, Y., KOYANAGI-AOI, M., HAYASHI, K., YAMANAKA, S., SAITOU, M. & LEE, JEANNIE T. 2013. Tsix RNA and the Germline Factor, PRDM14, Link X Reactivation and Stem Cell Reprogramming. *Molecular Cell,* 52**,** 805-818.

PEHRSON, J. R., CHANGOLKAR, L. N., COSTANZI, C. & LEU, N. A. 2014. Mice without MacroH2A Histone Variants. *Molecular and Cellular Biology,* 34**,** 4523-4533.

PEHRSON, J. R., COSTANZI, C. & DHARIA, C. 1997. Developmental and tissue expression patterns of histone macroH2A1 subtypes. *Journal of cellular biochemistry,* 65**,** 107-113.

PERCHARDE, M., LIN, C.-J., YIN, Y., GUAN, J., PEIXOTO, G. A., BULUT-KARSLIOGLU, A., BIECHELE, S., HUANG, B., SHEN, X. & RAMALHO-SANTOS, M. 2018. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell,* 174**,** 391-405. e19.

PEREZ, M. F. & LEHNER, B. 2019. Intergenerational and transgenerational epigenetic inheritance in animals. *Nature cell biology,* 21**,** 143-151.

PETRUSSA, L., VAN DE VELDE, H. & DE RYCKE, M. 2016. Similar kinetics for 5-methylcytosine and 5-hydroxymethylcytosine during human preimplantation development in vitro. *Molecular Reproduction and Development,* 83**,** 594-605.

PEZIC, D., MANAKOV, S. A., SACHIDANANDAM, R. & ARAVIN, A. A. 2014. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes & Development,* 28**,** 1410-1428.

PICCOLO, FRANCESCO M., BAGCI, H., BROWN, KAREN E., LANDEIRA, D., SOZA-RIED, J., FEYTOUT, A., MOOIJMAN, D., HAJKOVA, P., LEITCH, HARRY G., TADA, T., KRIAUCIONIS, S., DAWLATY, MEELAD M., JAENISCH, R., MERKENSCHLAGER, M. & FISHER, AMANDA G. 2013. Different Roles for Tet1 and Tet2 Proteins in Reprogramming-Mediated Erasure of Imprints Induced by EGC Fusion. *Molecular Cell,* 49**,** 1023-1033.

PICELLI, S., FARIDANI, O. R., BJÖRKLUND, Å. K., WINBERG, G., SAGASSER, S. & SANDBERG, R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols,* 9**,** 171-181.

PINHEIRO, I. & HEARD, E. 2017. X chromosome inactivation: new players in the initiation of gene silencing. *F1000Research,* 6**,** F1000 Faculty Rev-344.

PIROUZ, M., PILARSKI, S. & KESSEL, M. 2013. A Critical Function of Mad2l2 in Primordial Germ Cell Development of Mice. *PLOS Genetics,* 9**,** e1003712.

PLATH, K., FANG, J., MLYNARCZYK-EVANS, S. K., CAO, R., WORRINGER, K. A., WANG, H., DE LA CRUZ, C. C., OTTE, A. P., PANNING, B. & ZHANG, Y. 2003. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science,* 300**,** 131-135.

PLATT, R. N., 2ND, VANDEWEGE, M. W. & RAY, D. A. 2018. Mammalian transposable elements and their impacts on genome evolution. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology,* 26**,** 25-43.

PLONGTHONGKUM, N., DIEP, D. H. & ZHANG, K. 2014. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet,* 15**,** 647-661.

PODRINI, C., KOFFAS, A., CHOKSHI, S., VINCIGUERRA, M., LELLIOTT, C. J., WHITE, J. K., ADISSU, H. A., WILLIAMS, R. & GRECO, A. 2015. MacroH2A1 isoforms are associated with epigenetic markers for activation of lipogenic genes in fat-induced steatosis. *The FASEB Journal,* 29**,** 1676-1687.

PONNALURI, V. K. C., EHRLICH, K. C., ZHANG, G., LACEY, M., JOHNSTON, D., PRADHAN, S. & EHRLICH, M. 2017. Association of 5-hydroxymethylation and 5-methylation of DNA cytosine with tissue-specific gene expression. *Epigenetics,* 12**,** 123-138.

POSAVEC MARJANOVIĆ, M., HURTADO-BAGÈS, S., LASSI, M., VALERO, V., MALINVERNI, R., DELAGE, H., NAVARRO, M., CORUJO, D., GUBEROVIC, I., DOUET, J., GAMA-PEREZ, P., GARCIA-ROVES, P. M., AHEL, I., LADURNER, A. G., YANES, O., BOUVET, P., SUELVES, M., TEPERINO, R., POSPISILIK, J. A. & BUSCHBECK, M. 2017. MacroH2A1.1 regulates mitochondrial respiration by limiting nuclear NAD+ consumption. *Nature Structural &Amp; Molecular Biology,* 24**,** 902.

PRIKRYLOVA, T., ROBERTSON, J., FERRUCCI, F., KONORSKA, D., AANES, H., MANAF, A., ZHANG, B., VÅGBØ, C. B., KUŚNIERCZYK, A., GILLJAM, K. M., LØVKVAM-KØSTER, C., OTTERLEI, M., DAHL, J. A., ENSERINK, J., KLUNGLAND, A. & ROBERTSON, A. B. 2019. 5-hydroxymethylcytosine Marks Mammalian Origins Acting as a Barrier to Replication. *Scientific Reports,* 9**,** 11065.

PROKOPUK, L., STRINGER, J. M., HOGG, K., ELGASS, K. D. & WESTERN, P. S. 2017. PRC2 is required for extensive reorganization of H3K27me3 during epigenetic reprogramming in mouse fetal germ cells. *Epigenetics & Chromatin,* 10**,** 7.

QIU, X., MAO, Q., TANG, Y., WANG, L., CHAWLA, R., PLINER, H. A. & TRAPNELL, C. 2017. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods,* 14**,** 979-982.

QUARTA, C., SCHNEIDER, R. & TSCHÖP, MATTHIAS H. 2016. Epigenetic ON/OFF Switches for Obesity. *Cell,* 164**,** 341-342.

QUENNEVILLE, S., VERDE, G., CORSINOTTI, A., KAPOPOULOU, A., JAKOBSSON, J., OFFNER, S., BAGLIVO, I., PEDONE, PAOLO V., GRIMALDI, G., RICCIO, A. & TRONO, D. 2011. In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. *Molecular Cell,* 44**,** 361-372.

RAKYAN, V. K., BLEWITT, M. E., DRUKER, R., PREIS, J. I. & WHITELAW, E. 2002. Metastable epialleles in mammals. *Trends in Genetics,* 18**,** 348-351.

RAMOS-IBEAS, P., SANG, F., ZHU, Q., TANG, W. W. C., WITHEY, S., KLISCH, D., WOOD, L., LOOSE, M., SURANI, M. A. & ALBERIO, R. 2019. Pluripotency and X chromosome dynamics revealed in pig pre-gastrulating embryos by single cell analysis. *Nat Commun,* 10**,** 500.

RASSOULZADEGAN, M., GRANDJEAN, V., GOUNON, P., VINCENT, S., GILLOT, I. & CUZIN, F. 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature,* 441**,** 469-474.

Repeatmasker.org. 2020. Repeatmasker Genomic Datasets. [online] Available at: <http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html> [Accessed 7 September 2020].

REDDINGTON, J. P., PERRICONE, S. M., NESTOR, C. E., REICHMANN, J., YOUNGSON, N. A., SUZUKI, M., REINHARDT, D., DUNICAN, D. S., PRENDERGAST, J. G., MJOSENG, H., RAMSAHOYE, B. H., WHITELAW, E., GREALLY, J. M., ADAMS, I. R., BICKMORE, W. A. & MEEHAN, R. R. 2013. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biology,* 14**,** R25.

REUTER, M., BERNINGER, P., CHUMA, S., SHAH, H., HOSOKAWA, M., FUNAYA, C., ANTONY, C., SACHIDANANDAM, R. & PILLAI, R. S. 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature,* 480**,** 264-267.

REUTER, M., CHUMA, S., TANAKA, T., FRANZ, T., STARK, A. & PILLAI, R. S. 2009. Loss of the Mili-interacting Tudor domain–containing protein-1 activates transposons and alters the Mili-

associated small RNA profile. *Nature structural & molecular biology,* 16**,** 639.

REZNIK, B., CINCOTTA, S. A., JASZCZAK, R. G., MATEO, L. J., SHEN, J., CAO, M., BASKIN, L., YE, P., AN, W. & LAIRD, D. J. 2019. Heterogeneity of transposon expression and activation of the repressive network in human fetal germ cells. *Development,* 146**,** dev171157.

RICHLER, C., DHARA, S. K. & WAHRMAN, J. 2000. Histone macroH2A1.2 is concentrated in the XY compartment of mammalian male meiotic nuclei. *Cytogenetic and Genome Research,* 89**,** 118-120.

RIVERA-CASAS, C., GONZALEZ-ROMERO, R., CHEEMA, M. S., AUSIÓ, J. & EIRÍN-LÓPEZ, J. M. 2016. The characterization of macroH2A beyond vertebrates supports an ancestral origin and conserved role for histone variants in chromatin. *Epigenetics,* 11**,** 415-425.

ROBBEZ-MASSON, L., TIE, C. H., CONDE, L., TUNBAK, H., HUSOVSKY, C., TCHASOVNIKAROVA, I. A., TIMMS, R. T., HERRERO, J., LEHNER, P. J. & ROWE, H. M. 2018. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome research,* 28**,** 836-845.

ROBERT, C., KAPETANOVIC, R., BERALDI, D., WATSON, M., ARCHIBALD, A. L. & HUME, D. A. 2015. Identification and annotation of conserved promoters and macrophage-expressed genes in the pig genome. *BMC Genomics,* 16**,** 970.

RODRÍGUEZ-AGUILERA, J. R., ECSEDI, S., GOLDSMITH, C., CROS, M.-P., DOMÍNGUEZ-LÓPEZ, M., GUERRERO-CELIS, N., PÉREZ-CABEZA DE VACA, R., CHEMIN, I., RECILLAS-TARGA, F., CHAGOYA DE SÁNCHEZ, V. & HERNÁNDEZ-VARGAS, H. 2020. Genome-wide 5-hydroxymethylcytosine (5hmC) emerges at early stage of in vitro differentiation of a putative hepatocyte progenitor. *Scientific Reports,* 10**,** 7822.

ROSE, N. R. & KLOSE, R. J. 2014. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1839**,** 1362-1372.

ROTHBART, S. B., DICKSON, B. M., ONG, M. S., KRAJEWSKI, K., HOULISTON, S., KIREEV, D. B., ARROWSMITH, C. H. & STRAHL, B. D. 2013. Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. *Genes & Development,* 27**,** 1288-1298.

RUIZ, P. D. & GAMBLE, M. J. 2018. MacroH2A1 chromatin specification requires its docking domain and acetylation of H2B lysine 20. *Nature Communications,* 9**,** 5143.

RUNYAN, C., SCHAIBLE, K., MOLYNEAUX, K., WANG, Z., LEVIN, L. & WYLIE, C. 2006. Steel factor controls midline cell death of primordial germ cells and is essential for their normal proliferation and migration. *Development,* 133**,** 4861-4869.

RUSH, M., APPANAH, R., LEE, S., LAM, L. L., GOYAL, P. & LORINCZ, M. C. 2009. Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. *Epigenetics,* 4**,** 404-414.

RUZOV, A., TSENKINA, Y., SERIO, A., DUDNAKOVA, T., FLETCHER, J., BAI, Y., CHEBOTAREVA, T., PELLS, S., HANNOUN, Z., SULLIVAN, G., CHANDRAN, S., HAY, D. C., BRADLEY, M., WILMUT, I. & DE SOUSA, P. 2011. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res,* 21**,** 1332-1342.

SACHS, M., ONODERA, C., BLASCHKE, K., EBATA, K., SONG, J. & RAMALHO-SANTOS, M. 2013. Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline InVivo. *Cell Reports,* 3**,** 1777-1784.

SADO, T. & SAKAGUCHI, T. 2013. Species-specific differences in X chromosome inactivation in mammals. *Reproduction,* 146**,** R131-R139.

SAITOU, M., BARTON, S. C. & SURANI, M. A. 2002. A molecular programme for the specification of germ cell fate in mice. *Nature,* 418**,** 293-300.

SAITOU, M., KAGIWADA, S. & KURIMOTO, K. 2012. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development,* 139**,** 15-31.

SAITOU, M. & YAMAJI, M. 2012. Primordial germ cells in mice. *Cold Spring Harbor Perspectives in Biology,* 4**,** a008375.

SAKAI, Y., NAKAMURA, T., OKAMOTO, I., GYOBU-MOTANI, S., OHTA, H., YABUTA, Y., TSUKIYAMA, T., IWATANI, C., TSUCHIYA, H., EMA, M., MORIZANE, A., TAKAHASHI, J., YAMAMOTO, T. & SAITOU, M. 2019. Induction of the germ cell fate from pluripotent stem cells in cynomolgus monkeys†. *Biology of Reproduction,* 102**,** 620-638.

SAKASHITA, A., KAWABATA, Y., JINCHO, Y., TAJIMA, S., KUMAMOTO, S., KOBAYASHI, H., MATSUI, Y. & KONO, T. 2015. Sex Specification and Heterogeneity of Primordial Germ Cells in Mice. *PLoS ONE,* 10**,** e0144836.

SALVAING, J., AGUIRRE-LAVIN, T., BOULESTEIX, C., LEHMANN, G., DEBEY, P. & BEAUJEAN, N. 2012. 5-Methylcytosine and 5-hydroxymethylcytosine spatiotemporal profiles in the mouse zygote. *PloS one,* 7**,** e38156.

SANGRITHI, M. N., ROYO, H., MAHADEVAIAH, S. K., OJARIKRE, O., BHAW, L., SESAY, A., PETERS, A. H. F. M., STADLER, M. & TURNER, J. M. A. 2017. Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. *Developmental Cell,* 40**,** 289-301.e3.

SANTORO, S. W. & DULAC, C. 2015. Histone variants and cellular plasticity. *Trends in Genetics,* 31**,** 516-527.

SASAKI, K., NAKAMURA, T., OKAMOTO, I., YABUTA, Y., IWATANI, C., TSUCHIYA, H., SEITA, Y., NAKAMURA, S., SHIRAKI, N., TAKAKUWA, T., YAMAMOTO, T. & SAITOU, M. 2016. The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Developmental Cell,* 39**,** 169-185.

SASAKI, K., YOKOBAYASHI, S., NAKAMURA, T., OKAMOTO, I., YABUTA, Y., KURIMOTO, K., OHTA, H., MORITOKI, Y., IWATANI, C., TSUCHIYA, H., NAKAMURA, S., SEKIGUCHI, K., SAKUMA, T., YAMAMOTO, T., MORI, T., WOLTJEN, K.,

NAKAGAWA, M., YAMAMOTO, T., TAKAHASHI, K., YAMANAKA, S. & SAITOU, M. 2015. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell,* 17**,** 178-194.

SAUNDERS, P., TURNER, J., RUGGIU, M., TAGGART, M., BURGOYNE, P., ELLIOTT, D. & COOKE, H. 2003. Absence of mDazl produces a final block on germ cell development at meiosis. *Reproduction,* 126**,** 589-597.

SAXE, J. P., CHEN, M., ZHAO, H. & LIN, H. 2013. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *The EMBO journal,* 32**,** 1869-1885.

SAXONOV, S., BERG, P. & BRUTLAG, D. L. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 1412-1417.

SAYKALI, B., MATHIAH, N., NAHABOO, W., RACU, M.-L., HAMMOU, L., DEFRANCE, M. & MIGEOTTE, I. 2019. Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo. *eLife,* 8**,** e42434.

SCHACHTSCHNEIDER, K. M., MADSEN, O., PARK, C., RUND, L. A., GROENEN, M. A. M. & SCHOOK, L. B. 2015. Adult porcine genome-wide DNA methylation patterns support pigs as a biomedical model. *BMC Genomics,* 16**,** 743.

SCHMITGES, F. W., PRUSTY, A. B., FATY, M., STÜTZER, A., LINGARAJU, G. M., AIWAZIAN, J., SACK, R., HESS, D., LI, L. & ZHOU, S. 2011. Histone methylation by PRC2 is inhibited by active chromatin marks. *Molecular cell,* 42**,** 330-341.

SCHNEIDER, D. T., SCHUSTER, A. E., FRITSCH, M. K., HU, J., OLSON, T., LAUER, S., GÖBEL, U. & PERLMAN, E. J. 2001. Multipoint imprinting analysis indicates a common precursor cell for gonadal and nongonadal pediatric germ cell tumors. *Cancer research,* 61**,** 7268-7276.

SCHROEDER, D. I., JAYASHANKAR, K., DOUGLAS, K. C., THIRKILL, T. L., YORK, D., DICKINSON, P. J., WILLIAMS, L. E., SAMOLLOW, P. B., ROSS, P. J., BANNASCH, D. L., DOUGLAS, G. C. & LASALLE, J. M. 2015. Early Developmental and Evolutionary Origins of Gene Body DNA Methylation Patterns in Mammalian Placentas. *PLOS Genetics,* 11**,** e1005442.

SCHUETTENGRUBER, B., CHOURROUT, D., VERVOORT, M., LEBLANC, B. & CAVALLI, G. 2007. Genome Regulation by Polycomb and Trithorax Proteins. *Cell,* 128**,** 735-745.

SCHULZ, E. G. & HEARD, E. 2013. Role and control of X chromosome dosage in mammalian development. *Current Opinion in Genetics & Development,* 23**,** 109-115.

SCIALDONE, A., NATARAJAN, K. N., SARAIVA, L. R., PROSERPIO, V., TEICHMANN, S. A., STEGLE, O., MARIONI, J. C. & BUETTNER, F. 2015. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods,* 85**,** 54-61.

SEISENBERGER, S., ANDREWS, S., KRUEGER, F., ARAND, J., WALTER, J., SANTOS, F., POPP, C., THIENPONT, B., DEAN, W. & REIK, W.

2012. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular Cell,* 48**,** 849-862.

SEKI, Y., HAYASHI, K., ITOH, K., MIZUGAKI, M., SAITOU, M. & 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.

SEKI, Y., YAMAJI, M., YABUTA, Y., SANO, M., SHIGETA, M., MATSUI, Y., SAGA, Y., TACHIBANA, M., SHINKAI, Y. & SAITOU, M. 2007. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development,* 134.

SELLERS, Z. P., SCHNEIDER, G., MAJ, M. & RATAJCZAK, M. Z. 2018. Analysis of the Paternally-Imprinted DLK1–MEG3 and IGF2–H19 Tandem Gene Loci in NT2 Embryonal Carcinoma Cells Identifies DLK1 as a Potential Therapeutic Target. *Stem Cell Reviews and Reports,* 14**,** 823-836.

SEMBON, S., SUZUKI, S., FUCHIMOTO, D., IWAMOTO, M., KAWARASAKI, T. & ONISHI, A. 2008. Sex identification of pigs using polymerase chain reaction amplification of the amelogenin gene. *Zygote,* 16**,** 327-32.

SENFT, A. D., BIKOFF, E. K., ROBERTSON, E. J. & COSTELLO, I. 2019. Genetic dissection of Nodal and Bmp signalling requirements during primordial germ cell development in mouse. *Nature Communications,* 10**,** 1089.

SENFT, A. D., COSTELLO, I., KING, H. W., MOULD, A. W., BIKOFF, E. K. & ROBERTSON, E. J. 2018. Combinatorial Smad2/3 Activities Downstream of Nodal Signaling Maintain Embryonic/Extra-Embryonic Cell Identities during Lineage Priming. *Cell Reports,* 24**,** 1977-1985.e7.

SHAN, Y., LIANG, Z., XING, Q., ZHANG, T., WANG, B., TIAN, S., HUANG, W., ZHANG, Y., YAO, J., ZHU, Y., HUANG, K., LIU, Y., WANG, X., CHEN, Q., ZHANG, J., SHANG, B., LI, S., SHI, X., LIAO, B., ZHANG, C., LAI, K., ZHONG, X., SHU, X., WANG, J., YAO, H., CHEN, J., PEI, D. & PAN, G. 2017. PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. *Nature Communications,* 8**,** 672.

SHARIF, J., SHINKAI, Y. & KOSEKI, H. 2013. Is there a role for endogenous retroviruses to mediate long-term adaptive phenotypic response upon environmental inputs? *Philosophical Transactions of the Royal Society B: Biological Sciences,* 368**,** 20110340.

SHARMA, U., SUN, F., REICHHOLF, B., HERZOG, V., AMERES, S. & RANDO, O. 2017. Small RNAs are trafficked from the epididymis to developing mammalian sperm. *bioRxiv*.

SHEA, JEREMY M., SERRA, RYAN W., CARONE, BENJAMIN R., SHULHA, HENNADY P., KUCUKURAL, A., ZILLER, MICHAEL J., VALLASTER, MARKUS P., GU, H., TAPPER, ANDREW R., GARDNER, PAUL D., MEISSNER, A., GARBER, M. & RANDO, OLIVER J. 2015. Genetic and Epigenetic Variation, but Not Diet, Shape the Sperm Methylome. *Developmental Cell,* 35**,** 750-758.

SHEEDFAR, F., VERMEER, M., PAZIENZA, V., VILLARROYA, J., RAPPA, F., CAPPELLO, F., MAZZOCCOLI, G., VILLARROYA, F., VAN DER MOLEN, H., HOFKER, M. H., KOONEN, D. P. & VINCIGUERRA, M. 2015. Genetic ablation of macrohistone H2A1 leads to increased leanness, glucose tolerance and energy expenditure in mice fed a high-fat diet. *International Journal of Obesity,* 39**,** 331-338.

SHEN, L., INOUE, A., HE, J., LIU, Y., LU, F. & 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.

SHEN, L., WU, H., DIEP, D., YAMAGUCHI, S., D'ALESSIO, ANA C., FUNG, H.-L., ZHANG, K. & ZHANG, Y. 2013. Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. *Cell,* 153**,** 692-706.

SHIRANE, K., KURIMOTO, K., YABUTA, Y., YAMAJI, M., SATOH, J., ITO, S., WATANABE, A., HAYASHI, K., SAITOU, M. & SASAKI, H. 2016. Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. *Developmental Cell,* 39**,** 87-103.

SHOJI, M., TANAKA, T., HOSOKAWA, M., REUTER, M., STARK, A., KATO, Y., KONDOH, G., OKAWA, K., CHUJO, T. & SUZUKI, T. 2009. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Developmental cell,* 17**,** 775-787.

SIKLENKA, K., ERKEK, S., GODMANN, M., LAMBROT, R., MCGRAW, S., LAFLEUR, C., COHEN, T., XIA, J., SUDERMAN, M., HALLETT, M., TRASLER, J., PETERS, A. H. F. M. & KIMMINS, S. 2015. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science,* 350.

SKINNER, B. M., SARGENT, C. A., CHURCHER, C., HUNT, T., HERRERO, J., LOVELAND, J. E., DUNN, M., LOUZADA, S., FU, B., CHOW, W., GILBERT, J., AUSTIN-GUEST, S., BEAL, K., CARVALHO-SILVA, D., CHENG, W., GORDON, D., GRAFHAM, D., HARDY, M., HARLEY, J., HAUSER, H., HOWDEN, P., HOWE, K., LACHANI, K., ELLIS, P. J. I., KELLY, D., KERRY, G., KERWIN, J., NG, B. L., THREADGOLD, G., WILEMAN, T., WOOD, J. M. D., YANG, F., HARROW, J., AFFARA, N. A. & TYLER-SMITH, C. 2016. The pig X and Y Chromosomes: structure, sequence, and evolution. *Genome research,* 26**,** 130-139.

SKINNER, M. K. 2007. Endocrine disruptors and epigenetic transgenerational disease etiology. *Pediatric research,* 61**,** 48R-50R.

SKINNER, M. K., HAQUE, C. G.-B. M., NILSSON, E., BHANDARI, R. & MCCARREY, J. R. 2013. Environmentally Induced Transgenerational Epigenetic Reprogramming of Primordial Germ Cells and the Subsequent Germ Line. *PLoS ONE,* 8**,** e66318.

SKINNER, M. K., NILSSON, E., SADLER-RIGGLEMAN, I., BECK, D., BEN MAAMAR, M. & MCCARREY, J. R. 2019. Transgenerational sperm DNA methylation epimutation developmental origins following ancestral vinclozolin exposure. *Epigenetics,* 14**,** 721-739.

SLENTER, D. N., KUTMON, M., HANSPERS, K., RIUTTA, A., WINDSOR, J., NUNES, N., MÉLIUS, J., CIRILLO, E., COORT, S. L., DIGLES, D., EHRHART, F., GIESBERTZ, P., KALAFATI, M., MARTENS, M., MILLER, R., NISHIDA, K., RIESWIJK, L., WAAGMEESTER, A., EIJSSEN, L. M. T., EVELO, C. T., PICO, A. R. & WILLIGHAGEN, E. L. 2018. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research,* 46**,** D661-D667.

SMALLWOOD, S. A., TOMIZAWA, S.-I., KRUEGER, F., RUF, N., CARLI, N., SEGONDS-PICHON, A., SATO, S., HATA, K., ANDREWS, S. R. & KELSEY, G. 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet,* 43**,** 811-814.

SMITH, Z. D., CHAN, M. M., HUMM, K. C., KARNIK, R., MEKHOUBAD, S., REGEV, A., EGGAN, K. & MEISSNER, A. 2014. DNA methylation dynamics of the human preimplantation embryo. *Nature,* 511**,** 611-615.

SMITH, Z. D., CHAN, M. M., MIKKELSEN, T. S., GU, H., GNIRKE, A., REGEV, A. & MEISSNER, A. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature,* 484**,** 339-344.

SOARES, E. & ZHOU, H. 2018. Master regulatory role of p63 in epidermal development and disease. *Cellular and molecular life sciences : CMLS,* 75**,** 1179-1190.

SOH, Y. Q. S., ALFÖLDI, J., PYNTIKOVA, T., BROWN, L. G., GRAVES, T., MINX, P. J., FULTON, R. S., KREMITZKI, C., KOUTSEVA, N., MUELLER, J. L., ROZEN, S., HUGHES, J. F., OWENS, E., WOMACK, J. E., MURPHY, W. J., CAO, Q., DE JONG, P., WARREN, W. C., WILSON, R. K., SKALETSKY, H. & PAGE, D. C. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell,* 159**,** 800-813.

SOIBAM, B., BENHAM, A., KIM, J., WENG, K.-C., YANG, L., XU, X., ROBERTSON, M., AZARES, A., COONEY, A. J., SCHWARTZ, R. J. & LIU, Y. 2015. Genome-Wide Identification of MESP1 Targets Demonstrates Primary Regulation Over Mesendoderm Gene Activity. *STEM CELLS,* 33**,** 3254-3265.

SONG, Q., DECATO, B., HONG, E. E., ZHOU, M., FANG, F., QU, J., GARVIN, T., KESSLER, M., ZHOU, J. & SMITH, A. D. 2013. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One,* 8**,** e81148.

SOSA, E., CHEN, D., ROJAS, E. J., HENNEBOLD, J. D., PETERS, K. A., WU, Z., LAM, T. N., MITCHELL, J. M., SUKHWANI, M. & TAILOR, R. C. 2018. Differentiation of primate primordial germ cell-like cells following transplantation into the adult gonadal niche. *Nature communications,* 9**,** 1-13.

SPILLER, C. M. & BOWLES, J. 2015. Sex determination in mammalian germ cells. *Asian Journal of Andrology,* 17**,** 427-432.

SPRUIJT, CORNELIA G., GNERLICH, F., SMITS, ARNE H., PFAFFENEDER, T., JANSEN, PASCAL W. T. C., BAUER, C., MÜNZEL, M., WAGNER, M., MÜLLER, M., KHAN, F., EBERL, H.

C., MENSINGA, A., BRINKMAN, ARIE B., LEPHIKOV, K., MÜLLER, U., WALTER, J., BOELENS, R., VAN INGEN, H., LEONHARDT, H., CARELL, T. & VERMEULEN, M. 2013. Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell,* 152**,** 1146-1159.

STALLOCK, J., MOLYNEAUX, K., SCHAIBLE, K., KNUDSON, C. M. & WYLIE, C. 2003. The pro-apoptotic gene Bax is required for the death of ectopic primordial germ cells during their migration in the mouse embryo. *Development,* 130**,** 6589-6597.

STATHAM, A. L., ROBINSON, M. D., SONG, J. Z., COOLEN, M. W., STIRZAKER, C. & CLARK, S. J. 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome research,* 22**,** 1120-1127.

STEMMLER, M. P., ECCLES, R. L., BRABLETZ, S. & BRABLETZ, T. 2019. Non-redundant functions of EMT transcription factors. *Nature Cell Biology,* 21**,** 102-112.

STRINGER, J. M. & WESTERN, P. S. 2019. A step toward making human oocytes. *Nature Biotechnology,* 37**,** 24-25.

STROGANTSEV, R., KRUEGER, F., YAMAZAWA, K., SHI, H., GOULD, P., GOLDMAN-ROBERTS, M., MCEWEN, K., SUN, B., PEDERSEN, R. & FERGUSON-SMITH, A. C. 2015. Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biology,* 16**,** 112.

STROME, S. & UPDIKE, D. 2015. Specifying and protecting germ cell fate. *Nat Rev Mol Cell Biol,* 16**,** 406-416.

STROUD, H., FENG, S., MOREY KINNEY, S., PRADHAN, S. & JACOBSEN, S. E. 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biology,* 12**,** R54-R54.

STRYJEWSKA, A., DRIES, R., PIETERS, T., VERSTAPPEN, G., CONIDI, A., CODDENS, K., FRANCIS, A., UMANS, L., VAN IJCKEN, W. F. J., BERX, G., VAN GRUNSVEN, L. A., GROSVELD, F. G., GOOSSENS, S., HAIGH, J. J. & HUYLEBROECK, D. 2017. Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. *Stem cells (Dayton, Ohio),* 35**,** 611-625.

STUART, T., BUTLER, A., HOFFMAN, P., HAFEMEISTER, C., PAPALEXI, E., MAUCK, W. M., 3RD, HAO, Y., STOECKIUS, M., SMIBERT, P. & SATIJA, R. 2019. Comprehensive Integration of Single-Cell Data. *Cell,* 177**,** 1888-1902 e21.

SUGAWA, F., ARAÚZO‐BRAVO, M. J., YOON, J., KIM, K. P., ARAMAKI, S., WU, G., STEHLING, M., PSATHAKI, O. E., HÜBNER, K. & SCHÖLER, H. R. 2015. Human primordial germ cell commitment in vitro associates with a unique PRDM14 expression profile. *The EMBO Journal,* 34**,** 1009-1024.

SUGIMOTO, M. & ABE, K. 2007. X Chromosome Reactivation Initiates in Nascent Primordial Germ Cells in Mice. *PLOS Genetics,* 3**,** e116.

SULTANA, T., VAN ESSEN, D., SIOL, O., BAILLY-BECHET, M., PHILIPPE, C., ZINE EL AABIDINE, A., PIOGER, L., NIGUMANN, P., SACCANI, S., ANDRAU, J.-C., GILBERT, N. & CRISTOFARI, G.

2019. The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Molecular Cell,* 74**,** 555-570.e7.

SUN, Z. & BERNSTEIN, E. 2019. Histone variant macroH2A: from chromatin deposition to molecular function. *Essays In Biochemistry,* 63**,** 59-74.

SYBIRNA, A., TANG, W. W. C., PIERSON SMELA, M., DIETMANN, S., GRUHN, W. H., BROSH, R. & SURANI, M. A. 2020. A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. *Nature Communications,* 11**,** 1282.

TACHIBANA, M., MA, H., SPARMAN, M., LEE, H., RAMSEY, C., WOODWARD, J., SRITANAUDOMCHAI, H., MASTERSON, K., WOLFF, E. & JIA, Y. 2012. X-chromosome inactivation in monkey embryos and pluripotent stem cells. *Developmental Biology,* 371**,** 146-155.

TACHIBANA, M., SUGIMOTO, K., NOZAKI, M., UEDA, J., OHTA, T., OHKI, M., FUKUDA, M., TAKEDA, N., NIIDA, H., KATO, H. & SHINKAI, Y. 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & Development,* 16**,** 1779-1791.

TACHIBANA, M., UEDA, J., FUKUDA, M., TAKEDA, N., OHTA, T., IWANARI, H., SAKIHAMA, T., KODAMA, T., HAMAKUBO, T. & SHINKAI, Y. 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes & Development,* 19**,** 815-826.

TADA, M., TAKAHAMA, Y., ABE, K., NAKATSUJI, N. & TADA, T. 2001. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Current Biology,* 11**,** 1553-1558.

TAKADA, S., BEREZIKOV, E., CHOI, Y. L., YAMASHITA, Y. & MANO, H. 2009. Potential role of miR-29b in modulation of Dnmt3a and Dnmt3b expression in primordial germ cells of female mouse embryos. *RNA,* 15**,** 1507-1514.

TALBOT, N. C., REXROAD, C. E., PURSEL, V. G., POWELL, A. M. & NEL, N. D. 1993. Culturing the epiblast cells of the pig blastocyst. *In Vitro Cellular & Developmental Biology-Animal,* 29**,** 543-554.

TAM, P. P. L. & ZHOU, S. X. 1996. The Allocation of Epiblast Cells to Ectodermal and Germ-Line Lineages Is Influenced by the Position of the Cells in the Gastrulating Mouse Embryo. *Developmental Biology,* 178**,** 124-132.

TAN, FREDERICK E. & YEO, GENE W. 2016. Blurred Boundaries: The RNA Binding Protein Lin28A Is Also an Epigenetic Regulator. *Molecular Cell,* 61**,** 1-2.

TAN, S., TAN, H. T. & CHUNG, M. C. M. 2008. Membrane proteins and membrane proteomics. *PROTEOMICS,* 8**,** 3924-3932.

TANAKA, S. S., TOYOOKA, Y., AKASU, R., KATOH-FUKUI, Y., NAKAHARA, Y., SUZUKI, R., YOKOYAMA, M. & NOCE, T. 2000. The mouse homolog of Drosophila Vasa is required for the development of male germ cells. *Genes & Development,* 14**,** 841-853.

TANG, WALFRED W., DIETMANN, S., IRIE, N., LEITCH, HARRY G., FLOROS, VASILEIOS I., BRADSHAW, CHARLES R., HACKETT,

JAMIE A., CHINNERY, PATRICK F. & SURANI, M A. 2015. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell,* 161**,** 1453-1467.

TANG, W. W. C., KOBAYASHI, T., IRIE, N., DIETMANN, S. & SURANI, M. A. 2016. Specification and epigenetic programming of the human germ line. *Nat Rev Genet,* 17**,** 585-600.

TCHIEU, J., ZIMMER, B., FATTAHI, F., AMIN, S., ZELTNER, N., CHEN, S. & STUDER, L. 2017. A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. *Cell stem cell,* 21**,** 399-410.e7.

TIMINSZKY, G., TILL, S., HASSA, P. O., HOTHORN, M., KUSTATSCHER, G., NIJMEIJER, B., COLOMBELLI, J., ALTMEYER, M., STELZER, E. H. K., SCHEFFZEK, K., HOTTIGER, M. O. & LADURNER, A. G. 2009. A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nature Structural &Amp; Molecular Biology,* 16**,** 923.

TIROSH, I., IZAR, B., PRAKADAN, S. M., WADSWORTH, M. H., TREACY, D., TROMBETTA, J. J., ROTEM, A., RODMAN, C., LIAN, C. & MURPHY, G. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science,* 352**,** 189-196.

TISCHLER, J., GRUHN, W. H., REID, J., ALLGEYER, E., BUETTNER, F., MARR, C., THEIS, F., SIMONS, B. D., WERNISCH, L. & SURANI, M. A. 2019. Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate. *The EMBO Journal,* 38**,** e99518.

TITUS, L., HATCH, E. E., DRAKE, K. M., PARKER, S. E., HYER, M., PALMER, J. R., STROHSNITTER, W. C., ADAM, E., HERBST, A. L., HUO, D., HOOVER, R. N. & TROISI, R. 2019. Reproductive and hormone-related outcomes in women whose mothers were exposed in utero to diethylstilbestrol (DES): A report from the US National Cancer Institute DES Third Generation Study. *Reproductive Toxicology,* 84**,** 32-38.

TOMIZAWA, S.-I., KOBAYASHI, H., WATANABE, T., ANDREWS, S., HATA, K., KELSEY, G. & SASAKI, H. 2011. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development,* 138**,** 811-820.

TRAN, N. D., KISSNER, M., SUBRAMANYAM, D., PARCHEM, R. J., LAIRD, D. J. & BLELLOCH, R. H. 2016. A miR-372/let-7 Axis Regulates Human Germ Versus Somatic Cell Fates. *STEM CELLS,* 34**,** 1985-1991.

TRIANTAPHYLLOPOULOS, K. A., IKONOMOPOULOS, I. & BANNISTER, A. J. 2016. Epigenetics and inheritance of phenotype variation in livestock. *Epigenetics & Chromatin,* 9**,** 31.

TSUI, V. & CRISMANI, W. 2019. The Fanconi Anemia Pathway and Fertility. *Trends in Genetics,* 35**,** 199-214.

TSUKADA, Y.-I., AKIYAMA, T. & NAKAYAMA, K. I. 2015. Maternal TET3 is dispensable for embryonic development but is required for neonatal growth. *Scientific Reports,* 5**,** 15876.

TURNER, J. M. A., BURGOYNE, P. S. & SINGH, P. B. 2001. M31 and macroH2A1.2 colocalise at the pseudoautosomal region during mouse meiosis. *Journal of Cell Science,* 114**,** 3367-3375.

TYSER, R. C. V., MAHAMMADOV, E., NAKANOH, S., VALLIER, L., SCIALDONE, A. & SRINIVAS, S. 2020. A spatially resolved single cell atlas of human gastrulation. *bioRxiv***,** 2020.07.21.213512.

VALDEZ MAGAÑA, G., RODRÍGUEZ, A., ZHANG, H., WEBB, R. & ALBERIO, R. 2014. Paracrine effects of embryo-derived FGF4 and BMP4 during pig trophoblast elongation. *Developmental Biology,* 387**,** 15-27.

VALLOT, C., PATRAT, C., COLLIER, A. J., HURET, C., CASANOVA, M., LIYAKAT ALI, T. M., TOSOLINI, M., FRYDMAN, N., HEARD, E., RUGG-GUNN, P. J. & ROUGEULLE, C. 2017. XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell,* 20**,** 102-111.

VAN DE WERKEN, C., VAN DER HEIJDEN, G. W., ELEVELD, C., TEEUWSSEN, M., ALBERT, M., BAARENDS, W. M., LAVEN, J. S. E., PETERS, A. H. F. M. & BAART, E. B. 2014. Paternal heterochromatin formation in human embryos is H3K9/HP1 directed and primed by sperm-derived histone modifications. 5**,** 5868.

VAN DEN BRINK, S. C., ALEMANY, A., VAN BATENBURG, V., MORIS, N., BLOTENBURG, M., VIVIÉ, J., BAILLIE-JOHNSON, P., NICHOLS, J., SONNEN, K. F., MARTINEZ ARIAS, A. & VAN OUDENAARDEN, A. 2020. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature,* 582**,** 405-409.

VAN MIERLO, G., VEENSTRA, G. J. C., VERMEULEN, M. & MARKS, H. 2019. The Complexity of PRC2 Subcomplexes. *Trends in Cell Biology,* 29**,** 660-671.

VASSOLER, F. M., WHITE, S. L., SCHMIDT, H. D., SADRI-VAKILI, G. & PIERCE, R. C. 2013. Epigenetic Inheritance of a Cocaine Resistance Phenotype. *Nature neuroscience,* 16**,** 42-47.

VEJLSTED, M., DU, Y., VAJTA, G. & MADDOX-HYTTEL, P. 2006a. Post-hatching development of the porcine and bovine embryo—defining criteria for expected development in vivo and in vitro. *Theriogenology,* 65**,** 153-165.

VEJLSTED, M., OFFENBERG, H., THORUP, F. & MADDOX-HYTTEL, P. 2006b. Confinement and clearance of OCT4 in the porcine embryo at stereomicroscopically defined stages around gastrulation. *Molecular Reproduction and Development,* 73**,** 709-718.

VELAND, N., LU, Y., HARDIKAR, S., GADDIS, S., ZENG, Y., LIU, B., ESTECIO, M. R., TAKATA, Y., LIN, K., TOMIDA, M. W., SHEN, J., SAHA, D., GOWHER, H., ZHAO, H. & CHEN, T. 2019. DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic acids research,* 47**,** 152-167.

VELLA, P., SCELFO, A., JAMMULA, S., CHIACCHIERA, F., WILLIAMS, K., CUOMO, A., ROBERTO, A., CHRISTENSEN, J., BONALDI, T., HELIN, K. & PASINI, D. 2013. Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Molecular Cell,* 49**,** 645-656.

VERMA, N., PAN, H., DORÉ, L. C., SHUKLA, A., LI, Q. V., PELHAM-WEBB, B., TEIJEIRO, V., GONZÁLEZ, F., KRIVTSOV, A., CHANG, C.-J., PAPAPETROU, E. P., HE, C., ELEMENTO, O. & HUANGFU, D. 2018. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nature genetics,* 50**,** 83-95.

VÉRTESY, Á., ARINDRARTO, W., ROOST, M. S., REINIUS, B., TORRENS-JUANEDA, V., BIALECKA, M., MOUSTAKAS, I., ARIYUREK, Y., KUIJK, E., MEI, H., SANDBERG, R., VAN OUDENAARDEN, A. & CHUVA DE SOUSA LOPES, S. M. 2018. Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells. *Nature Communications,* 9**,** 1873.

VICKERS, M. H. 2014. Early Life Nutrition, Epigenetics and Programming of Later Life Disease. *Nutrients,* 6**,** 2165-2178.

VICKOVIC, S., ERASLAN, G., SALMÉN, F., KLUGHAMMER, J., STENBECK, L., SCHAPIRO, D., ÄIJÖ, T., BONNEAU, R., BERGENSTRÅHLE, L., NAVARRO, J. F., GOULD, J., GRIFFIN, G. K., BORG, Å., RONAGHI, M., FRISÉN, J., LUNDEBERG, J., REGEV, A. & STÅHL, P. L. 2019. High-definition spatial transcriptomics for in situ tissue profiling. *Nature methods,* 16**,** 987-990.

VIRE, E., BRENNER, C., DEPLUS, R., BLANCHON, L., FRAGA, M., DIDELOT, C., MOREY, L., VAN EYNDE, A., BERNARD, D., VANDERWINDEN, J.-M., BOLLEN, M., ESTELLER, M., DI CROCE, L., DE LAUNOIT, Y. & FUKS, F. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature,* 439**,** 871-874.

VOIGT, P., LEROY, G., DRURY, W. J., ZEE, B. M., SON, J., BECK, D. B., YOUNG, N. L., GARCIA, B. A. & REINBERG, D. 2012. Asymmetrically modified nucleosomes. *Cell,* 151**,** 181-193.

VON MEYENN, F., BERRENS, REBECCA V., ANDREWS, S., SANTOS, F., COLLIER, AMANDA J., KRUEGER, F., OSORNO, R., DEAN, W., RUGG-GUNN, PETER J. & REIK, W. 2016. Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. *Developmental Cell,* 39**,** 104-115.

VON MEYENN, F. & REIK, W. 2015. Forget the Parents: Epigenetic Reprogramming in Human Germ Cells. *Cell,* 161**,** 1248-1251.

WAINER-KATSIR, K. & LINIAL, M. 2016. Single cell expression data reveal human genes that escape X-chromosome inactivation. *bioRxiv***,** 079830.

WALTER, M., TEISSANDIER, A., PÉREZ-PALACIOS, R. & BOURC'HIS, D. 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife,* 5**,** e11418.

WALTERS, E. M., AGCA, Y., GANJAM, V. & EVANS, T. 2011. Animal models got you puzzled?: think pig. *Annals of the New York Academy of Sciences,* 1245**,** 63-64.

WAN, D., LIU, C., SUN, Y., WANG, W., HUANG, K. & ZHENG, L. 2017a. MacroH2A1.1 cooperates with EZH2 to promote adipogenesis by

regulating Wnt signaling. *Journal of Molecular Cell Biology,* 9**,** 325-337.

WAN, G., FIELDS, B., SPRACKLIN, G., PHILLIPS, C. & KENNEDY, S. 2017b. Transgenerational Epigenetic Inheritance Factors Localize to Spatially and Temporally Ordered Liquid Droplet Assemblages. *bioRxiv*.

WANG, H., XIANG, J., ZHANG, W., LI, J., WEI, Q., ZHONG, L., OUYANG, H. & HAN, J. 2016. Induction of germ cell-like cells from porcine induced pluripotent stem cells. *Scientific reports,* 6**,** 1-13.

WANG, J., ZHANG, M., ZHANG, Y., KOU, Z., HAN, Z., CHEN, D.-Y., SUN, Q.-Y. & GAO, S. 2010. The Histone Demethylase JMJD2C Is Stage-Specifically Expressed in Preimplantation Mouse Embryos and Is Required for Embryonic Development1. *Biology of Reproduction,* 82**,** 105-111.

WANG, L., ZHANG, J., DUAN, J., GAO, X., ZHU, W., LU, X., YANG, L., ZHANG, J., LI, G., CI, W., LI, W., ZHOU, Q., ALURU, N., TANG, F., HE, C., HUANG, X. & LIU, J. 2014. Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell,* 157**,** 979-991.

WANG, T., XU, C., LIU, Y., FAN, K., LI, Z., SUN, X., OUYANG, H., ZHANG, X., ZHANG, J., LI, Y., MACKENZIE, F., MIN, J. & TU, X. 2012. Crystal Structure of the Human SUV39H1 Chromodomain and Its Recognition of Histone H3K9me2/3. *PLOS ONE,* 7**,** e52977.

WATANABE, N., MII, S., ASAI, N., ASAI, M., NIIMI, K., USHIDA, K., KATO, T., ENOMOTO, A., ISHII, H. & TAKAHASHI, M. 2013. The REV7 subunit of DNA polymerase ζ is essential for primordial germ cell maintenance in the mouse. *Journal of Biological Chemistry,* 288**,** 10459-10471.

WATANABE, T., CHUMA, S., YAMAMOTO, Y., KURAMOCHI-MIYAGAWA, S., TOTOKI, Y., TOYODA, A., HOKI, Y., FUJIYAMA, A., SHIBATA, T. & SADO, T. 2011. MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Developmental cell,* 20**,** 364-375.

WEBER, A. R., KRAWCZYK, C., ROBERTSON, A. B., KUŚNIERCZYK, A., VÅGBØ, C. B., SCHUERMANN, D., KLUNGLAND, A. & SCHÄR, P. 2016. Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications,* 7**,** 10806.

WEBER, M., HELLMANN, I., STADLER, M. B., RAMOS, L., PAABO, S., REBHAN, M. & SCHUBELER, D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet,* 39**,** 457-466.

WEI, X., GUO, J., LI, Q., JIA, Q., JING, Q., LI, Y., ZHOU, B., CHEN, J., GAO, S., ZHANG, X., JIA, M., NIU, C., YANG, W., ZHI, X., WANG, X., YU, D., BAI, L., WANG, L., NA, J., ZOU, Y., ZHANG, J., ZHANG, S. & MENG, D. 2019. Bach1 regulates self-renewal and impedes mesendodermal differentiation of human embryonic stem cells. *Science Advances,* 5**,** eaau7887.

WEI, Y., YANG, C.-R., WEI, Y.-P., ZHAO, Z.-A., HOU, Y., SCHATTEN, H. & SUN, Q.-Y. 2014. Paternally induced transgenerational inheritance

of susceptibility to diabetes in mammals. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 1873-1878.

WEIGEL, D. & COLOT, V. 2012. Epialleles in plant evolution. *Genome Biology,* 13**,** 249-249.

WEN, J., ZENG, Y., FANG, Z., GU, J., GE, L., TANG, F., QU, Z., HU, J., CUI, Y., ZHANG, K., WANG, J., LI, S., SUN, Y. & JIN, Y. 2017. Single-cell analysis reveals lineage segregation in early post-implantation mouse embryos. *The Journal of biological chemistry,* 292**,** 9840-9854.

WEN, Y.-A., XIONG, X., ZAYTSEVA, Y. Y., NAPIER, D. L., VALLEE, E., LI, A. T., WANG, C., WEISS, H. L., EVERS, B. M. & GAO, T. 2018. Downregulation of SREBP inhibits tumor growth and initiation by altering cellular metabolism in colon cancer. *Cell Death & Disease,* 9**,** 265.

WEST, J. A., VISWANATHAN, S. R., YABUUCHI, A., CUNNIFF, K., TAKEUCHI, A., PARK, I.-H., SERO, J. E., ZHU, H., PEREZ-ATAYDE, A., FRAZIER, A. L., SURANI, M. A. & DALEY, G. Q. 2009. A role for Lin28 in primordial germ cell development and germ cell malignancy. *Nature,* 460**,** 909-913.

WHITELAW, E. 2015. Disputing Lamarckian epigenetic inheritance in mammals. *Genome Biology,* 16**,** 60.

WILLIAMS, K., CHRISTENSEN, J., PEDERSEN, M. T., JOHANSEN, J. V., CLOOS, P. A. C., RAPPSILBER, J. & HELIN, K. 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature,* 473**,** 343-348.

WOLF, G., DE IACO, A., SUN, M.-A., BRUNO, M., TINKHAM, M., HOANG, D., MITRA, A., RALLS, S., TRONO, D. & MACFARLAN, T. S. 2020. KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *eLife,* 9**,** e56337.

WOLF, X. A., SERUP, P. & HYTTEL, P. 2011. Three-dimensional localisation of NANOG, OCT4, and E-CADHERIN in porcine pre- and peri-implantation embryos. *Dev Dyn,* 240**,** 204-10.

WOSSIDLO, M., NAKAMURA, T., LEPIKHOV, K., MARQUES, C. J., ZAKHARTCHENKO, V., BOIANI, M., ARAND, J., NAKANO, T., REIK, W. & WALTER, J. 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun,* 2**,** 241.

WU, H., COSKUN, V., TAO, J., XIE, W., GE, W., YOSHIKAWA, K., LI, E., ZHANG, Y. & SUN, Y. E. 2010. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science,* 329**,** 444-448.

WU, H., D'ALESSIO, A. C., ITO, S., XIA, K., WANG, Z., CUI, K., ZHAO, K., EVE SUN, Y. & ZHANG, Y. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature,* 473**,** 389-393.

WU, H. & ZHANG, Y. 2011. Tet1 and 5-hydroxymethylation: A genome-wide view in mouse embryonic stem cells. *Cell Cycle,* 10**,** 2428-2436.

WU, Q., ZHANG, L., SU, P., LEI, X., LIU, X., WANG, H., LU, L., BAI, Y., XIONG, T., LI, D., ZHU, Z., DUAN, E., JIANG, E., FENG, S., HAN, M., XU, Y., WANG, F. & ZHOU, J. 2015. MSX2 mediates entry of

human pluripotent stem cells into mesendoderm by simultaneously suppressing SOX2 and activating NODAL signaling. *Cell research,* 25**,** 1314-1332.

WU, X. & ZHANG, Y. 2017. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics,* 18**,** 517-534.

XIOL, J., CORA, E., KOGLGRUBER, R., CHUMA, S., SUBRAMANIAN, S., HOSOKAWA, M., REUTER, M., YANG, Z., BERNINGER, P. & PALENCIA, A. 2012. A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing. *Molecular cell,* 47**,** 970-979.

XU, C., XU, Y., GURSOY-YUZUGULLU, O. & PRICE, B. D. 2012. The histone variant macroH2A1.1 is recruited to DSBs through a mechanism involving PARP1. *FEBS letters,* 586**,** 3920-3925.

XU, Y., WU, F., TAN, L., KONG, L., XIONG, L., DENG, J., BARBERA, A. J., ZHENG, L., ZHANG, H., HUANG, S., MIN, J., NICHOLSON, T., CHEN, T., XU, G., SHI, Y., ZHANG, K. & SHI, YUJIANG G. 2011. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Molecular Cell,* 42**,** 451-464.

YAMAGUCHI, S., HONG, K., LIU, R., INOUE, A., SHEN, L., ZHANG, K. & ZHANG, Y. 2013a. Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Research,* 23**,** 329-339.

YAMAGUCHI, S., HONG, K., LIU, R., SHEN, L., INOUE, A., DIEP, D., ZHANG, K. & ZHANG, Y. 2012. Tet1 controls meiosis by regulating meiotic gene expression. *Nature,* 492**,** 443-447.

YAMAGUCHI, S., KIMURA, H., TADA, M., NAKATSUJI, N. & TADA, T. 2005. Nanog expression in mouse germ cell development. *Gene Expression Patterns,* 5**,** 639-646.

YAMAGUCHI, S., KURIMOTO, K., YABUTA, Y., SASAKI, H., NAKATSUJI, N., SAITOU, M. & TADA, T. 2009. Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells. *Development,* 136**,** 4011-4020.

YAMAGUCHI, S., SHEN, L., LIU, Y., SENDLER, D. & 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. & SAITOU, M. 2008. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nature genetics,* 40**,** 1016.

YAMASHIRO, C., SASAKI, K., YABUTA, Y., KOJIMA, Y., NAKAMURA, T., OKAMOTO, I., YOKOBAYASHI, S., MURASE, Y., ISHIKURA, Y., SHIRANE, K., SASAKI, H., YAMAMOTO, T. & SAITOU, M. 2018. Generation of human oogonia from induced pluripotent stem cells in vitro. *Science,* 362**,** 356-360.

YANG, H., LIN, H., XU, H., ZHANG, L., CHENG, L., WEN, B., SHOU, J., GUAN, K., XIONG, Y. & YE, D. 2014. TET-catalyzed 5-methylcytosine hydroxylation is dynamically regulated by metabolites. *Cell Research,* 24**,** 1017-1020.

YANG, J., RYAN, D. J., WANG, W., TSANG, J. C.-H., LAN, G., MASAKI, H., GAO, X., ANTUNES, L., YU, Y., ZHU, Z., WANG, J., KOLODZIEJCZYK, A. A., CAMPOS, L. S., WANG, C., YANG, F., ZHONG, Z., FU, B., ECKERSLEY-MASLIN, M. A., WOODS, M., TANAKA, Y., CHEN, X., WILKINSON, A. C., BUSSELL, J., WHITE, J., RAMIREZ-SOLIS, R., REIK, W., GÖTTGENS, B., TEICHMANN, S. A., TAM, P. P. L., NAKAUCHI, H., ZOU, X., LU, L. & LIU, P. 2017a. Establishment of mouse expanded potential stem cells. *Nature,* 550**,** 393.

YANG, P., WANG, Y. & MACFARLAN, T. S. 2017b. The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. *Trends in Genetics,* 33**,** 871-881.

YANG, S., DING, S., HE, S., HE, L., GAO, K., PENG, S. & SHUAI, C. 2019. Differentiation of primordial germ cells from premature ovarian insufficiency-derived induced pluripotent stem cells. *Stem Cell Research & Therapy,* 10**,** 156.

YANG, Z., CHEN, K.-M., PANDEY, R. R., HOMOLKA, D., REUTER, M., JANEIRO, B. K. R., SACHIDANANDAM, R., FAUVARQUE, M.-O., MCCARTHY, A. A. & PILLAI, R. S. 2016. PIWI slicing and EXD1 drive biogenesis of nuclear piRNAs from cytosolic targets of the mouse piRNA pathway. *Molecular cell,* 61**,** 138-152.

YILDIRIM, O., LI, R., HUNG, J.-H., CHEN, P. B., DONG, X., EE, L.-S., WENG, Z., RANDO, O. J. & FAZZIO, T. G. 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell,* 147**,** 1498-1510.

YODER, J. A., WALSH, C. P. & BESTOR, T. H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics,* 13**,** 335-340.

YOKOBAYASHI, S., OKITA, K., NAKAGAWA, M., NAKAMURA, T., YABUTA, Y., YAMAMOTO, T. & SAITOU, M. 2017. Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate†. *Biology of Reproduction,* 96**,** 1154-1166.

YOSHIDA, M., KAJIKAWA, E., KUROKAWA, D., TOKUNAGA, T., ONISHI, A., YONEMURA, S., KOBAYASHI, K., KIYONARI, H. & AIZAWA, S. 2016. Conserved and divergent expression patterns of markers of axial development in eutherian mammals. *Dev Dyn,* 245**,** 67-86.

YOSHIMURA, T., TOYODA, S., KURAMOCHI-MIYAGAWA, S., MIYAZAKI, T., MIYAZAKI, S., TASHIRO, F., YAMATO, E., NAKANO, T. & MIYAZAKI, J.-I. 2009. Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Developmental biology,* 335**,** 216-227.

YOUNG, N. L., DIMAGGIO, P. A., PLAZAS-MAYORCA, M. D., BALIBAN, R. C., FLOUDAS, C. A. & GARCIA, B. A. 2009. High Throughput Characterization of Combinatorial Histone Codes. *Molecular & Cellular Proteomics,* 8**,** 2266-2284.

YU, D., WANG, J., ZOU, H., FENG, T., CHEN, L., LI, J., QI, X., LI, Z., DUAN, X., XU, C., ZHANG, L., LONG, X., LAN, J., CHEN, C., WANG, C., XU, X., REN, J., ZHAO, Y., HU, X., LIAN, Z., MEN, H.,

PAN, D., LI, N., CAPECCHI, M. R., DU, X., ZHAO, Y. & WU, S. 2018a. Silencing of retrotransposon-derived imprinted gene RTL1 is the main cause for postimplantational failures in mammalian cloning. *Proceedings of the National Academy of Sciences,* 115**,** E11071-E11080.

YU, X.-X., LIU, Y.-H., LIU, X.-M., WANG, P.-C., LIU, S., MIAO, J.-K., DU, Z.-Q. & YANG, C.-X. 2018b. Ascorbic acid induces global epigenetic reprogramming to promote meiotic maturation and developmental competence of porcine oocytes. *Scientific Reports,* 8**,** 6132.

YUAN, S., OLIVER, D., SCHUSTER, A., ZHENG, H. & YAN, W. 2015. Breeding scheme and maternal small RNAs affect the efficiency of transgenerational inheritance of a paramutation in mice. *Scientific Reports,* 5**,** 9266.

YUNHAI, L. & C., S. J. 2015. MCM9 deficiency delays primordial germ cell proliferation independent of the ATM pathway. *genesis,* 53**,** 678-684.

ZELLER, P., PADEKEN, J., VAN SCHENDEL, R., KALCK, V., TIJSTERMAN, M. & GASSER, S. M. 2016. Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet,* 48**,** 1385-1395.

ZENG, Y., YAO, B., SHIN, J., LIN, L., KIM, N., SONG, Q., LIU, S., SU, Y., GUO, JUNJIE U., HUANG, L., WAN, J., WU, H., QIAN, J., CHENG, X., ZHU, H., MING, G.-L., JIN, P. & SONG, H. 2016. Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Molecular Cell,* 61**,** 153-160.

ZHANG, B., ZHENG, H., HUANG, B., LI, W., XIANG, Y., PENG, X., MING, J., WU, X., ZHANG, Y., XU, Q., LIU, W., KOU, X., ZHAO, Y., HE, W., LI, C., CHEN, B., LI, Y., WANG, Q., MA, J., YIN, Q., KEE, K., MENG, A., GAO, S., XU, F., NA, J. & XIE, W. 2016a. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature,* 537**,** 553-557.

ZHANG, J., ZHANG, M., ACAMPORA, D., VOJTEK, M., YUAN, D., SIMEONE, A. & CHAMBERS, I. 2018a. OTX2 restricts entry to the mouse germline. *Nature,* 562**,** 595-599.

ZHANG, M., LEITCH, H. G., TANG, W. W. C., FESTUCCIA, N., HALL-PONSELE, E., NICHOLS, J., SURANI, M. A., SMITH, A. & CHAMBERS, I. 2018b. Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. *Cell reports,* 22**,** 332-339.

ZHANG, S., CHEN, X., WANG, F., AN, X., TANG, B., ZHANG, X., SUN, L. & LI, Z. 2016b. Aberrant DNA methylation reprogramming in bovine SCNT preimplantation embryos. *Scientific Reports,* 6**,** 30345.

ZHANG, W., XIA, W., WANG, Q., TOWERS, A. J., CHEN, J., GAO, R., ZHANG, Y., YEN, C.-A., LEE, A. Y. & LI, Y. 2016c. Isoform switch of TET1 regulates DNA demethylation and mouse development. *Molecular cell,* 64**,** 1062-1073.

ZHANG, X., COZEN, A. E., LIU, Y., CHEN, Q. & LOWE, T. M. 2016d. Small RNA Modifications: Integral to Function and Disease. *Trends in molecular medicine*.

ZHANG, X., SU, J., JEONG, M., KO, M., HUANG, Y., PARK, H. J., GUZMAN, A., LEI, Y., HUANG, Y.-H., RAO, A., LI, W. &

GOODELL, M. A. 2016e. DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat Genet,* 48**,** 1014-1023.

ZHAO, L., SUN, M.-A., LI, Z., BAI, X., YU, M., WANG, M., LIANG, L., SHAO, X., ARNOVITZ, S. & WANG, Q. 2014. The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. *Genome research,* 24**,** 1296-1307.

ZHAO, Q., ZHANG, J., CHEN, R., WANG, L., LI, B., CHENG, H., DUAN, X., ZHU, H., WEI, W., LI, J., WU, Q., HAN, J.-D. J., YU, W., GAO, S., LI, G. & WONG, J. 2016. Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nature Communications,* 7**,** 12464.

ZHENG, H., HUANG, B., ZHANG, B., XIANG, Y., DU, Z., XU, Q., LI, Y., WANG, Q., MA, J., PENG, X., XU, F. & XIE, W. 2016. Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Molecular Cell,* 63**,** 1066-1079.

ZHENG, Y., XUE, X., SHAO, Y., WANG, S., ESFAHANI, S. N., LI, Z., MUNCIE, J. M., LAKINS, J. N., WEAVER, V. M., GUMUCIO, D. L. & FU, J. 2019. Controlled modelling of human epiblast and amnion development using stem cells. *Nature,* 573**,** 421-425.

ZHOU, W., LIANG, G., MOLLOY, P. L. & JONES, P. A. 2020. DNA methylation enables transposable element-driven genome expansion. *Proceedings of the National Academy of Sciences,* 117**,** 19359-19366.

ZHU, H., WANG, G. & QIAN, J. 2016. Transcription factors as readers and effectors of DNA methylation. *Nature Reviews Genetics,* 17**,** 551.

ZHU, P., GUO, H., REN, Y., HOU, Y., DONG, J., LI, R., LIAN, Y., FAN, X., HU, B., GAO, Y., WANG, X., WEI, Y., LIU, P., YAN, J., REN, X., YUAN, P., YUAN, Y., YAN, Z., WEN, L., YAN, L., QIAO, J. & TANG, F. 2018. Single-cell DNA methylome sequencing of human preimplantation embryos. *Nature Genetics,* 50**,** 12-19.

ZOU, H., YU, D., DU, X., WANG, J., CHEN, L., WANG, Y., XU, H., ZHAO, Y., ZHAO, S., PANG, Y., LIU, Y., HAO, H., ZHAO, X., DU, W., DAI, Y., LI, N., WU, S. & ZHU, H. 2019. No imprinted XIST expression in pigs: biallelic XIST expression in early embryos and random X inactivation in placentas. *Cellular and Molecular Life Sciences,* 76**,** 4525-4538.

ZYLICZ, J. J., DIETMANN, S., GÜNESDOGAN, U., HACKETT, J. A., COUGOT, D., LEE, C. & SURANI, M. A. 2015. Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *eLife,* 4**,** e09571.

Appendix

Appendix Table 1 List of primary antibodies and their working dilutions used in this project

	Host	Brand	ID	Dilution
PGC markers				
Sda/GM2	Mouse	Home-made	/	1:50-1:100
SOX17	Goat	R&D	AF1924	1:500-1:700
DAZL	Goat	Novusbio	NB100-2437	1:200
GPR50	Rabbit	Proteintech	19762-1-AP	1:100-1:200
DNA Repair Components				
UNG	Rabbit	Proteintech	12394-1-AP	1:300
PARP1	Rabbit	Proteintech	13371-1-AP	1:200
Epigenetic Markers and modifiers				
5hmC	Rabbit	Active Motif	39791	1:1500-1:2000
5mC	Mouse	Eurogentec	BI-MECY-0100	1:1500-1:2000
H3K27me3	Rabbit	Active Motif	39157	1:300
H2AFY	Mouse	R&D	MAB8318	1:100
EED	Rabbit	Proteintech	16818-1-AP	1:300
FACS of human PGC				
Alexa Fluor 488-conjugated anti-alkaline phosphatase (AP)	Mouse	BD Pharmingen	561495	1:100

PerCP-Cy5.5-conjugated anti-CD117	Mouse	BD Pharmingen	333950	1:100

Appendix Table 2 Sequencing statistics of the PBAT libraries

PigRef	Total Reads	Mapped Reads				Reads after Deduplication				Total CpG Sites
		Counts			Percentage	Counts			Percentage	
		PE (pbat)	R1 (pbat)	R2 (default)		PE (pbat)	R1 (pbat)	R2 (default)		
female_pgc1	706,566,896	264,461,670	71,566,960	56,888,793	55.61%	175,780,996	38,254,067	27,035,292	34.12%	59,318,274
female_pgc2	700,348,120	279,010,304	75,177,934	57,847,185	58.83%	197,493,700	41,611,372	27,356,243	38.05%	59,318,274
female_soma	650,868,108	244,820,062	69,457,217	52,101,976	56.29%	176,640,974	40,499,123	26,475,766	37.43%	59,318,274
male_pgc1	687,650,606	256,693,940	59,746,398	39,540,248	51.77%	194,738,178	40,864,740	25,816,539	38.02%	59,318,274
male_pgc2	670,897,138	263,151,760	49,382,055	38,821,361	52.37%	195,421,282	34,622,981	24,742,798	37.98%	59,318,274
male_soma	661,616,978	275,325,482	41,235,774	29,627,307	52.32%	198,559,132	28,096,368	19,344,930	37.18%	59,318,274

PigRef	Cytosine Coverage								Bisulfite Conversion Rate
	>=1x		>=3x		>=5x		>=10x		
	Counts	Percentage	Counts	Percentage	Counts	Percentage	Counts	Percentage	
female_pgc1	55,624,073	93.77%	46,200,315	77.89%	35,142,971	59.24%	13,558,696	22.86%	99.20%
female_pgc2	55,756,579	94.00%	47,032,097	79.29%	36,994,182	62.37%	16,945,242	28.57%	99.10%
female_soma	55,610,740	93.75%	45,697,267	77.04%	33,858,009	57.08%	11,602,690	19.56%	99.00%
male_pgc1	55,835,892	94.13%	47,330,603	79.79%	37,557,788	63.32%	17,622,753	29.71%	99.10%
male_pgc2	55,208,021	93.07%	45,234,503	76.26%	34,673,710	58.45%	15,485,676	26.11%	99.20%
male_soma	54,701,611	92.22%	43,234,818	72.89%	31,354,930	52.86%	11,161,879	18.82%	98.90%

Appendix Table 3 List of all biological samples used in this project

Description	Cell Type	Pooled/Single Embryo	Isolation	Number of Cells
scRNA-seq				
E14 Pig Embryo	PGC	pooled	FACS	17
	Soma	pooled	FACS	89
E31 Pig Embryo	PGC	pooled	FACS	22
PBAT				
Wk5 (E35) Female Pig Embryo 1	PGC	Single	FACS	5000
Wk5 (E35) Female Pig Embryo 2	PGC	Single	FACS	7778
Wk5 (E35) Female Pig Embryo 3	Soma	Single	FACS	5000
Wk5 (E35) Male Pig Embryo 1	PGC	Single	FACS	5139
	Soma	Single	FACS	5000
Wk5 (E35) Male Pig Embryo 2	PGC	Single	FACS	6760
LC-MS				
Pig Samples				
E11 pig Epiblast 1	Epiblast	pooled	Dissection	/
E11 pig Epiblast 2	Epiblast	pooled	Dissection	/
E11 pig Epiblast 3	Epiblast	pooled	Dissection	/
E14 Pig PGC 1	PGC	pooled	FACS	1120
E14 Pig PGC 2	PGC	pooled	FACS	660
E14 Pig PGC 3	PGC	pooled	FACS	841
E14 Pig Soma 1	Soma	pooled	FACS	5000
E14 Pig Soma 2	Soma	pooled	FACS	>5000
Wk6 Pig PGC 1	PGC	pooled	FACS	7790
Wk6 Pig PGC 2	PGC	pooled	FACS	8210
Wk6 Pig Soma 1	Soma	pooled	FACS	5000
Wk6 Pig Soma 2	Soma	pooled	FACS	5000
Human Samples				
hPGCLC rep1	PGC	/	FACS	1000
hPGCLC rep2	PGC	/	FACS	1000
hPGCLC rep3	PGC	/	FACS	500
soma rep1 (induction)	Soma	/	FACS	1000

soma rep2 (induction)	Soma	/	FACS	1000
Wk7 (E51) hPGC	PGC	Single	FACS	1722
Wk7 (E48) hPGC	PGC	Single	FACS	1302
Wk7 (E51) hSoma	Soma	Single	FACS	10000
Wk7 (E48) hSoma	Soma	Single	FACS	10000
hESC (H9) rep1	ESC	/	/	>2 million
hESC (H9) rep2	ESC	/	/	>1 million