Title

Primordial Germ Cell Differentiation in Vitro: A Model for Understanding Epigenetic Reprogramming and Genome-Wide DNA Demethylation in Mouse Primordial Germ Cells

Author

Vincent, John J.

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Primordial Germ Cell Differentiation in Vitro: A Model for Understanding Epigenetic Reprogramming and Genome-Wide DNA Demethylation in Mouse Primordial Germ Cells

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

John James Vincent

2013

ABSTRACT OF THE DISSERTATION

Primordial Germ Cell Differentiation in Vitro: A Model for Understanding Epigenetic Reprogramming and Genome-Wide DNA Demethylation in Mouse Primordial Germ Cells

by

John James Vincent

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2013

Professor Amander Therese Clark, Chair

Sperm and oocytes are terminally differentiated, sex-specific **germ cells**, which, upon fertilization will generate a new embryo and leads to species propagation by sexual reproduction. Though fated only to generate eggs or sperm, germ cells have the unique property to imbue zygotes with totipotent capacity, which facilitates the formation of all tissues the embryo will need to survive to adulthood. These characteristics are facilitated by germ cells' ability to pass on genetic information to the next generation, as well as their capacity to initiate genome-wide reorganization and removal of epigenetic information inherited by germ cells during embryogenesis. It is hypothesized that remodeling of this epigenetic information is essential to drive proper embryo development. While these events are not well understood, it is known that the events that underlie these unique properties are initiated in early development, shortly after

the germ line is established as a pool of **primordial germ cells (PGCs)**. Efforts to unravel these mechanisms that underlie totipotency in germ cells have been limited due to the inability to isolate, study, and manipulate PGCs. To overcome this obstacle, we hypothesized that PGCs cells could be differentiated from pluripotent embryonic stem cells, and that these cells would serve as a surrogate cell type for the study of PGC biology.

Establishing a new model of lineage differentiation from embryonic stem cells required the development of assays and criteria to rigorously test identity, developmental staging, and epigenetic progression to determine if an in vitro model is able to recapitulate features of endogenous PGCs. To accomplish this, we developed a scalable and transgene-free method to differentiate immature PGCs in vitro using the cell surface markers SSEA1 and cKit that are developmentally and epigenetically reminiscent of immature PGCs. We applied existing assays to validate PGC identity, and devised a new stringent assay based on genetic deletion of a known PGC determinant. We developed a single-cell gene expression methodology to compare gene expression signatures of in vitro derived PGCs and endogenous PGCs, and identified novel criteria to define PGC identity from early endogenous PGCs and in vitro-generated PGCs.

We next used in vitro PGC differentiation to investigate genome-wide DNA demethylation, one of the first epigenetic reprogramming events undertaken by early PGCs. By combining the scalability of this differentiation system with next generation methylation sequencing techniques, we generated the first DNA methylation maps of in vitro derived PGCs, and determined with sequence-specific information that DNA demethylation is genome-wide and likely to involve loss of DNA methylation as a consequence of cell division. We also investigated potentiators of

active DNA methylation loss, including the Tet proteins, and their roles in early PGC development.

Finally, we applied single cell gene expression technology to define developmental progression of human PGCs isolated from the gonads of fetuses from elective terminations. We identified a common progenitor stage of PGC development in the human fetal gonad. Furthermore, we adapted our single cell gene expression approaches to interrogate differentiation strategies in the generation of the common human PGC progenitor in vitro.

Together, we have developed a differentiation system to ask questions about epigenetic progression in early germ cells, and have utilized single cell gene expression technology and genomics to characterize seminal events in the epigenetic reprogramming of human and mouse germ cells.

The dissertation of John James Vincent is approved.

Feng Guo

Steven Erik Jacobsen

Harley I. Kornblum

April Dawn Pyle

Amander Therese Clark, Chair

University of California, Los Angeles

2013

DEDICATION

In memory of my grandfather, John Lienhard

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I have to thank my mentor, Amander, for the knowledge, guidance, and support that you have given me over than past five years of my training. Thank you for giving me the opportunity to achieve everything that we have together. Thank you for your time, effort, and unrelenting determination to push me and make me better because of it. I also want to thank my committee members: Steve, April, Feng, and Harley, for the great committee meetings, and discussions about my research. Thank you for the countless critiques, contributions and committee meetings that went into my development as a scientist. I also want to thank all of our collaborators on these projects: Matteo Pellegrini, Pao-Yang Chen, Suhua Feng for all their help with the genomics; Juehua Yu and Thuc Le, for mass spectrometry, reagents, and stimulating scientific discussion; and Michaela Patterson, Akanksha Chhabara, Ben Van Handel, and Bill Lowry, for our scientific discussions, collaborations, and generally being awesome neighbors.

To the Clark lab-you guys made coming in to work every day worth it. First, I have to acknowledge the postdocs. To Sofia-thank you for the support (both scientific and emotional), the coffee breaks, and for sharing in our successes over these years. You are a true friend. To Anne- you taught me everything you know, which allowed me to do everything that I was able to accomplish. You provided great insight and guidance into how to be a science while still retaining your sanity. Thank you, thank you, thank you. To the grad students: thanks for sharing both the failures and successes we have experienced together in the grueling process that is grad school. Ziwei-you have been an amazing colleague, collaborator, and friend for all of these years. I will truly miss working with you and your spirit. Danke ZZ! To Serena, Marisabel, Joe-thank you for being great friends. Lab would not be as fun without the stories, D-listed, and nostalgic ‘90s Pandora music stations. So many of the experiments that are

presented here would not have been even possible without the FACS expertise of Felicia and Codrea and Jessica Scholes, who performed every single sort with me over the past 5+ years. It is because of you that we were able to make these awesome discoveries.

I owe a lot of the success in the works presented in this dissertation to the undergraduates who volunteered to help fearlessly and unwaveringly. You guys pushed these projects forward and really made them your own in the ways you contributed to them. Your unwavering eagerness to learn reminds me why I like science. To Xian, Kevin, Sara-I have learned just as much about being a teacher, scientist, and mentor as I hope you've learned from me. Thank you for every contribution, big and small, that each of you have made to these studies. You will all make amazing scientists one day.

To my friends-you made grad school the amazing memory I know I'll cherish forever. To Sean, Gabe, Akanksha, Andrew, Ben, Michaela, Karina, Gabi, Lynn, Jenny, Faith, Jamie, Nicole, Maria-thank you for sharing the journey with me.

And last, but certainly not least, I thank my family--Mom, Jacob, Grandma, Lori, Kenny, James, and Ben--for your unwavering support of my endeavors. Even though we've been separated geographically, I know you support me and believe that I can accomplish anything. I hope this work is proof of that. I love you.

Chapter 2 is a copy of the accepted version of the manuscript "Vincent JJ, Li Z, Lee SA, Liu X, Etter MO, Gkountela S, Diaz-Perez SV, Lindgren AG, Clark AT. (2011). Single cell analysis facilitates staging of *Blimp1*-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS ONE.* PMID: 22194959" published in *PLoS ONE* in December 2011. It is reproduced here freely by author-retained copyright.

Chapter 3 is a version of the manuscript "Vincent JJ, Huang Y, Chen PY, Feng S, Calvopina JH, Nee K, Lee SA, Le T, Yoon AJ, Faull K, Fan G, Rao A, Jacobsen SE, Pellegrini M, Clark AT. (2013). Stage specific roles for Tet1/Tet2 in DNA demethylation in primordial germ cells" currently accepted for publication at *Cell Stem Cell.*

Chapter 4 is a modified version of the manuscript "Gkountela S, Li Z, Vincent JJ, Zhang KX, Pellegrini M, Clark AT. (2012). The ontogeny of cKIT+ human primordial germ cells: A resource for human germ line reprogramming, imprint erasure and *in vitro* differentiation. *Nature Cell Biology;* 15(1) pp.113-22*"* previously published in December 2012. It is reproduced here freely by author-retained copyright.

I thank the funding which supported my research and travel over the course of my dissertation: California Institute for Regenerative Medicine Grants T-00005 and TG2-01169, the Molecular Biology Institute for travel grant funding, and the UCLA Dissertation Year Fellowship.

VITA

2007	B.A., Biochemistry and Molecular Biology Boston University Boston, Massachusetts
2009-10	Teaching Assistant Molecular, Cell and Developmental Biology University of California, Los Angeles
2008-10	Award Recipient / Trainee California Institute for Regenerative Medicine (CIRM) Pre-doctoral Training Grant University of California, Los Angeles
2011	Award Recipient / Trainee Eli and Edythe Broad Center for Regenerative Medicine Pre-doctoral Training Grant University of California, Los Angeles
2012	Award Recipient Dissertation Year Fellowship University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Andken BB, Lim I, Benson G, Vincent JJ, Ferenc MT, Heinrich B, Jarzylo LA, Man HY, Deshler JO. (2007). 3'UTR-SIRF: A Database for Identifying Clusters of Short Interspersed Repeats in 3'-Untranslated Regions. *BMC Bioinformatics*. PMID: 17663765.

Vincent JJ, Clark AT. (2008). *MicroRNA Biogenesis is Essential for Tumorigenesis and the Emergence of Cancer Stem Cells.* UCLA Molecular Cell and Developmental Biology Retreat, Lake Arrowhead, CA.

Vincent JJ, Clark AT. (2009). *Characterizing the Role of MicroRNAs in Pluripotent Stem Cell Transformation.* International Society for Stem Cell Research 7th Meeting, Barcelona, Spain.

Vincent JJ, Li Z, Lindgren AG, Clark AT. (2009). *Establishing the Principles for in vitro Gametogenesis.* UCLA MBI Departmental Retreat, 2009, Lake Arrowhead, CA.

Vincent JJ, Clark AT. (2010). *Understanding Early Events in the Emerging Germ Line Using Pluripotent Stem Cells.* UCLA MCDB Departmental Retreat 2010, Lake Arrowhead, CA.

Ambartsumyan G, Gill RK, Perez SD, Conway D, Vincent JJ, Dalal Y, Clark AT. (2010). Centromere protein A dynamics in human pluripotent stem cell self-renewal, differentiation, and DNA damage. *Hum. Mol Genet*. PMID: 20650959.

Vincent JJ, Li Z, Oliveros M, Lindgren AG, Clark AT. (2010). *Three Dimensional Gonad Reconstructions Support Maturation of in vitro derived Primordial Germ Cells.* International Society for Stem Cell Research 8th Meeting, San Francisco, CA.

Vincent JJ, Li Z, Oliveros M, Lindgren AG, Clark AT. (2010). *Three Dimensional Gonad Reconstructions Support Maturation of in vitro derived Primordial Germ Cells.* 1st Annual Broad Tri-Institutional Retreat, 2010, Asilomar, CA.

Vincent JJ, Li Z, Oliveros M, Lindgren AG, Clark AT, (2010). *Three Dimensional Gonad Reconstructions Support Maturation of in vitro derived Primordial Germ Cells.* Cold Spring Harbor 2010 Meeting on Germ Cells, Cold Spring Harbor, NY.

Lindgren AG, Natsuhara K, Diaz-Perez SV, Vincent JJ, Jiao J, Wu H, Clark AT. (2010). *Loss of Pten causes tumor initiation after differentiation of murine pluripotent stem cells.* UCLA 6th Annual Stem Cell Conference, 2010, Los Angeles, CA.

Ambartsumyan G, Diaz-Perez S, Conway D, Vincent JJ, Clark AT. (2010). CENP-A *mRNA Reserves are Required in Transitioning from Pluripotency to Embryonic Lineage Differentiation.* UCLA 6th Annual Stem Cell Conference, 2010, Los Angeles, CA.

Vincent JJ, Li Z, Oliveros M, Lindgren AG, Clark AT. (2010). *Three Dimensional Gonad Reconstructions Support Maturation of in vitro derived Primordial Germ Cells.* California Institute of Regenerative Medicine 2nd Annual Grantee Meeting, 2010, San Francisco, CA.

Ambartsumyan G, Diaz-Perez SV, Conway D, Vincent JJ, Clark AT. (2010). CENP-A *mRNA Reserves are Required in Transitioning from Pluripotency to Embryonic Lineage Differentiation.* California Institute of Regenerative Medicine 2nd Annual Grantee Meeting, 2010, San Francisco, CA.

Vincent JJ, Li Z, Lee SA, Liu X, Etter MO, Gkountela S, Diaz-Perez SV, Lindgren AG, Clark AT. (2011). Direct targets of the master regulator *Blimp1* in primordial germ cells derived from mouse embryonic stem cells. UCLA 7th Annual Stem Cell Conference, Los Angeles, CA.

Vincent JJ, Li Z, Lee SA, Liu X, Etter MO, Gkountela S, Diaz-Perez SV, Lindgren AG, Clark AT. (2011). Single cell analysis facilitates staging of *Blimp1*-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS ONE.* PMID: 22194959.

Lindgren AG, Natsuhara K, Tian E, Vincent JJ, Li X, Jiao J, Wu H, Banerjee U, Clark AT. (2011). Loss of *Pten* causes tumor initiation from differentiated pluripotent cells due to failed repression of *Nanog. PLoS ONE.* PMID: 21304588.

Vincent JJ, Chen PY, Feng S, Le T, Jacobsen SE, Pellegrini M, Fan G, Clark AT. (2011). Oxidation of 5-methylcytosine as a Mechanism of Genome Wide DNA Demethylation in Mouse Primordial Germ Cells. UCLA Molecular Cell & Developmental Biology Retreat, Lake Arrowhead, CA.

Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M, Clark AT. (2012). The ontogeny of cKIT+ human primordial germ cells: A resource for human germ line reprogramming, imprint erasure and *in vitro* differentiation. *Nature Cell Biology.* PMID: 23242216.

Vincent JJ, Huang Y, Chen PY, Feng S, Calvopina JH, Nee K, Lee SA, Le T, Yoon AJ, Faull K, Fan G, Rao A, Jacobsen SE, Pellegrini M, Clark AT. (2013). Stage specific roles for Tet1 and Tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell,* in press.

CHAPTER 1

INTRODUCTION

An Overview of Critical Events in Early Germ Line Development

Gametes—oocytes and sperm—are terminally differentiated, haploid germ cells tasked with ensuring organisms are reproductively fit to generate offspring that have the capacity to develop into fertile adults (Figure 1-1). This requires germ cells to perform two key roles: to pass parental DNA on to the next generation, and to ensure that zygotes have the necessary epigenetic landscape to facilitate embryo development. Mammalian germ cells are specified via inductive signaling on pluripotent epiblast cells in post-implantation embryos to generate the founding pool of **primordial germ cells (PGCs)** [1-3]. To achieve successful development and imbue an organism with reproductive fitness, germ cells utilize transcriptional repression of somatic gene expression as well as dramatic plasticity in the remodeling of their epigenomes after they have been specified [2,4,5]. It is hypothesized that germ cells undertake this epigenetic reprogramming in order to confer epigenetic totipotency to offspring in the next generation [2,6]. However, the regulators of this crucial developmental process are not characterized or well understood.

Early PGCs are difficult to isolate and study. Furthermore, our ability to ask intrinsic questions abut PGC biology is hampered by the lack of a clear, faithful *in vitro* model to recapitulate mammalian germ line development. In this chapter, I will discuss the known mechanisms of PGC specification, development, maturation and survival. Second, I will discuss the known epigenetic events that occur during PGC ontogeny, and speculate on possible mechanisms PGCs may undertake to initiate drastic epigenetic remodeling, particularly as it pertains to DNA demethylation. Finally, I will discuss the current state of *in vitro* systems to model this early time frame of development, and identify key criteria necessary for development of a robust and faithful model to study PGC differentiation and early development.

1. PGC Specification

For the completion of embryogenesis by newly formed zygotes it is critical that genetic and epigenetic material passed on through the gametes can support the development of healthy, new individuals. This process is initiated early in embryogenesis, at embryonic day (E)7.25 in the mouse in the epiblast. During this period of early development, germ cells are referred to as **primordial germ cells (PGCs)** [2,4]. PGCs are specified by inductive paracrine signaling on epiblast cells from adjacent extraembryonic ectoderm and visceral endoderm cells and is mainly mediated by ligands *Bmp2, Bmp4,* and *Bmp8b* beginning at E6.0 [7-13]. By E6.25, the expression of ***Prdm1/Blimp1***, a transcriptional repressor essential for PGC development, becomes detectable within a subset of BMP-primed epiblast cells, and will go on to generate lineage-restricted PGCs by E7.25, when expression of ***Stella/Dppa3***, the earliest known marker of founder PGCs protein is apparent [2,4].

The importance of *Blimp1* in pre-gonadal PGC development

Blimp1 is a transcriptional repressor that contains a SET domain and Kruppel-type zinc fingers within its coding region, which confer DNA-binding activity to the protein[14]. In B cells, Blimp1 has been characterized for its role as a master regulator of terminal B cell differentiation into antibody-secreting plasma cells through direct repression of the existing mature B-cell genetic program [15-19]. In the embryo, global deletion of *Blimp1* is embryonic lethal by E10.5 due to placental defects [20,21]. PGC-specific deletion of *Blimp1* causes a striking depletion of Stella+ founder PGCs from E7.25 embryos, and haploinsufficiency experiments have shown that this requirement is dose-dependent (Vincent 2005, Ohinata 2005). Of note, *Blimp1* null PGC-like cells fail to repress somatic lineage genes, notably the homeobox genes *Hoxa1* and *Hoxb1* [4,21]. A closer examination of the effect of Blimp1 deletion on PGCs has been provided by Kurimoto

and colleagues, who performed a comprehensive microarray expression analyses performed on single PGCs and somatic neighbors [22]. This study demonstrated that *Blimp1* null PGC-like cells have aberrant expression of multiple somatic genes that are normally repressed in wild type PGCs, including genes involved in gastrulation, pattern specification, morphogenesis, cell cycle, and DNA methylation machinery. Therefore, Blimp1 drives PGC specification mainly through the repression of somatic gene programs, although some evidence has indicated that *Blimp1* may also induce expression of genes that promote PGC identity [22].

Blimp1 and Protein Arginine Methyltransferase 5 (Prmt5)

While Blimp1 has DNA-binding capacity conferred by its zinc fingers, there is no enzymatic activity associated with Blimp1. It has been hypothesized that Blimp1 interacts with effector molecules to mediate gene expression changes associated with cellular differentiation, and that Blimp1 confers gene specificity through recognition of a Blimp1 recognition motif within target genes. Indeed, studies from B-cells have identified a Blimp1 consensus binding sequence in target genes directly bound by Blimp1 during plasma cell differentiation of terminally differentiated B cells [23], which I have also observed in many of the Blimp1-regulated target somatic genes identified by Kurimoto and colleagues (Figure 1-2)[22].

The molecular mechanism by which Blimp1 facilitates development in PGCs is less clear. Blimp1 immunoprecipitates from somatic cells can direct methylation of the third arginine residue of the histone tails of H2A and H4. These marks are generally considered to be repressive marks, and have been hypothesized to be the major effectors of the repression of somatic lineages in PGCs [24]. Further studies found that Blimp1 interacts with Protein Arginine Methyltransferase 5 (Prmt5) *in vitro*, and that they are co-expressed in individual PGCs. Prmt5

directs H2A/H4R3 dimethylation [25], and is enriched at a Blimp1 consensus motif found within the *Dhx38* locus in chromatin from E10.5 genital ridges [24]. These results have led to a model whereby the interaction of Blimp1 with Prmt5 and key consensus sequences mediate somatic cell gene repression to facilitate PGC development. However, lack of cellular material has restricted efforts to clarify the role of *Blimp1* in directing PGC development, or that Blimp1 and Prmt5 physically interact in PGCs *in vitro* or *in vivo*.

DNA Methylation is Inherited from the Epiblast into PGCs

Though critical for PGC development, transcriptional establishment of the germ cell program via *Blimp1* is insufficient to drive germ cell development and generate gametes that can undergo fertilization and convey totipotency to offspring. Animals have evolved multiple mechanisms to establish the germ cells during embryogenesis. In lower organisms, including *Drosophila, C. elegans,* and zebrafish, germ cells are formed via a process called **preformation**, or the selective sequestration of maternal factors into newly dividing embryonic cells, and these cells are fated to become the source of gametes for the adult life of the organism. In contrast, mammalian PGCs are specified via **induction** from pluripotent epiblast cells, which have potential to become all cell lineages of the organism.

In order to facilitate embryo differentiation and lineage formation prior to PGC induction, epiblast cells undergo the massive addition of methyl groups to cytosine bases in the DNA, generating **5-methylcytosine (5mC)**. 5mC is deposited by the *de novo* DNA methyltransferases *Dnmt3a* and Dnmt3b [26]. The observation of the accumulation of 5mC in the epiblast has been appreciated for decades by methylation-sensitive DNA digestion studies of early cell populations [27]. Strikingly, crude preparations of PGC genomic DNA were found to have markedly

hypomethylated genomes compared to somatic and epiblast cell populations. These findings lead to a model where PGCs may escape de novo methylation in PGC-fated epiblast precursors. This model would imply that preformation of the germ line in mammals may act to protect germ line establishment, while the remaining epiblast cells form somatic lineages [27,28].

However, later studies of murine PGCs were able to show that this is not the case. Experiments using 5mC-specific antibodies found that, to the contrary, PGCs indeed inherit methylated DNA in their genomes from the parental epiblast cells up to E8.0[29]. Other techniques, including PCR of bisulfite-treated DNA and next-generation sequencing of PGC genomes and are markedly hypomethylated by E13.5 [29-32]. These findings lead to a model where PGCs have mechanisms to remove 5mC from the genome after methylation patterns have been established rather than shielded from methylation as proposed [27].

Why is the removal of 5mC from PGC genomes necessary? 5mC is essential for lineage specialization [33], and epiblast cells must undergo differentiation to generate the specialized cell types of an organism. The removal of 5mC from PGCs during this period has been proposed to be essential for establishing the germ line ground state, marked by near complete loss of DNA methylation from PGCs[6,31,32,34]. The erasure of 5mC from PGC genomes is thought to be necessary to imbue totipotent capacity to PGCs[5], which is essential for the generation of future offspring[29,35,36]. The process of removal of DNA methylation occurs in two phases, beginning at E8.0.

2. PGC Development, E8.0-E10.5

After specification, PGCs reside in the allantois until E8.0. At this time point, PGCs begin to initiate a migratory phase through the developing embryo to reach the future gonad. Embryological studies of this period have shown that PGCs depend heavily on survival signals and migratory cues [37]. Accompanying this time, however, PGCs are also undergoing intrinsic changes to their epigenetic landscape including DNA methylation and histone modifications. These events are temporally associated with the initiation of PGC migration, and involve changes to DNA methylation and histone modifications [29].

PGC Migration

PGCs initiate migration heterogeneously by E8.0, beginning with the entrance of PGCs in the hindgut endoderm adjacent to the allantois until E9.5. After E9.5, PGCs migrate bilaterally from the embryo midline, and begin to pool in the genital ridges beginning at E10.5-E11.5.

Embryological studies have shown that PGCs are dependent upon signaling interactions between the surrounding soma and PGCs. One of the primary signaling pathways essential for this period of PGC life is the receptor tyrosine kinase **c-Kit,** expressed on the surface of PGCs[38]. cKit is expressed by PGCs continually throughout migration, and its ligand, **Stem Cell Factor (SCF)**, is expressed by surrounding somatic cells and is thought to provide a spatio-temporal niche signal during migration [39,40]. Mutation in either SCF or cKit results in sterility, with reduced number of migratory PGCs observed. cKit signaling is thought to promote survival for PGCs, as supplementation of SCF to PGCs cultured *in vitro* promotes proliferation [41]. In vivo, it has been shown that SCF/cKit signaling functions to suppress Bax-mediated apoptosis in PGCs thus acting as a survival signal [40]. cKit signaling appears to only affect migration rate of PGCs and not direction, as PGCs of SCF null embryos still migrate in the proper direction [39].

In addition to the roles of SCF and cKit in PGC survival, a role for chemoattractive signaling between PGCs and soma has also been identified, and is critical for attracting PGCs from the endoderm to the genital ridges. In particular, SDF1/CXCR4 signaling appears to be critical for attracting PGCs to move laterally from the midline to colonize the gonad. Stromal Derived Factor 1 (SDF1) is a ligand expressed by somatic cells of the genital ridge, and its receptor CXCR4 is expressed in PGCs, and is conserved down to zebrafish [42]. In mice, loss of either SDF1 or CXCR4 results in very few PGCs reaching the genital ridges. Furthermore, ectopic expression of SDF1 induces ectopic mislocalization of PGCs to new embryonic locations [43,44].

Step 1 Epigenetic Reprogramming

In addition to the signaling milieu that mediates PGC migration and survival, PGCs also initiate massive epigenetic changes beginning at E8.0. Many of these changes, particularly DNA methylation, are associated with a remodeling of the epigenetic identity inherited from the parental epiblast. It has been hypothesized that this remodeling in early embryonic life is essential for reproductive fitness of an individual [6]. Many of the studies into this period of development are based on descriptive studies involving immunofluorescence and examination of methylation of specific interesting loci [29,30,35,45], and the molecular mechanisms that drive these massive epigenetic changes are largely uncharacterized. Nonetheless, these studies have demonstrated that changes to DNA methylation apparently occur in a discrete manner comprised of two phases. In the first phase, epigenetic remodeling involves a drastic but incomplete reduction in DNA methylation as well as remodeling of histone modifications from E8.0 to approximately E10.5, concomitant with PGC migration (Figure 1-3)[29]. Notably, methylation is maintained at imprinted genes despite global loss of methylated cytosine by

immunofluorescence. Collectively, the pre-gonadal epigenetic changes associated with pre-gonadal PGC development are termed **Reprogramming Step 1**. A second step of reprogramming occurs coincident with PGCs entering the gonad from >E10.5-E13.5, where further demethylation occurs at imprinted genes [30]. It has also been speculated that global DNA methylation may be further reduced during this time. Together, these events are termed **Reprogramming Step 2** (Figure 1-3)**.**

Rapid changes in histone modifications occur shortly after specification of PGCs from epiblast cells [29]. Dimethylation of the ninth lysine of histone H3 (H3K9m2), a canonical repressive histone mark associated with heterochromatin, is lost heterogeneously as early as E7.75 PGCs as they begin to exist the allantois and begin migration [45]. H3K9m2 loss is preceded by a rapid downregulation of *Glp*, the histone methyltransferase essential for depositing this mark, in lineage-restricted Stella+ PGCs shortly after specification at E7.75 [45]. Trimethylation of lysine 27 on Histone H3 (H3K27m3) then quickly becomes heavily deposited in PGC nuclei over the course of 12 hours, beginning at E8.5 and approaching maximal levels by E9.5, which are sustained through E11.5 [29,46]. The function of the changing histone modifications roughly over these three days in development are not understood, but it has been proposed that the exchange of H3K27m3 for H3K9m2 may facilitate demethylation by providing increased accessibility to the DNA, as H3K27m3 is considered a more plastic epigenetic mark that can be remodeled more readily [29]. Interestingly, loss of H3K9 and gain of H3K27m3 occurs concomitantly with a G2 arrest phase that PGCs undergo from E7.75-8.75. In this phase, very few PGCs enter S phase, and up to 60% of PGCs have been found to accumulate in G2 arrest [45]. In contrast, PGCs at E10.5 are actively cycling and proliferating, with no such block in the cell cycle. This has led to speculation that remodeling of PGC epigenomes may be regulated by

their ability to re-enter the cell cycle after appropriate epigenetic modifications have been made [45].

Of particular interest, one of the most striking events to occur during Step 1 Reprogramming is the genome-wide loss of 5mC bases from genomic DNA, which is initiated shortly after PGCs are specified from epiblast precursors. Descriptive studies using antibodies that recognize 5mC have shown that PGCs exhibit specific demethylation upon the initiation of Reprogramming I at E8.0-8.5. and is preceded by a transient loss of all Dnmt expression [29]. Even though *Dnmt1* (maintenance) and *Dnmt3a (de novo)* become re-expressed at E8.75, the global decrease in 5mC is maintained throughout PGC migration, and is never observed in the surrounding embryonic somatic cells [29].

3. Step 2 Reprogramming in the Gonad, >E10.5-E13.5

Upon gonadal colonization, PGCs undergo further changes to histones and subsequent locus-specific demethylation of imprinted genes and single copy genes [30]. PGCs retain reduced 5mC staining throughout migration, and then undergo an even further drop in 5mC signal around E11.5-E12.5, at the onset of Step 2 Reprogramming [29,47]. At this point, genome-wide methylation levels are as low as 16% and represent the most basal level of DNA methylation during PGC ontogeny [31]. Given that there are two discernible phases of DNA demethylation, and that IF studies indicate that the wave during Reprogramming 1 is genome-wide, with distinct mechanisms to protect distinct loci, it is possible that distinct mechanism regulate discrete phases of PGC demethylation.

Changes to histone modifications also accompany PGCs as they enter Step 2 reprogramming. At E11.5, rapid loss of linker histone H1 takes place and is associated with increased nuclear diameter, presumably due to the decondensation of chromatin (Hajkova 2008). Loss of H3K27m3, H2AR3m2 and H4R3m2 modifications also precedes loss of H2A.Z at E11.5 [24,46]. Given that the cell cycle of PGCs is approximately 16 hours, the striking depletion of robust histone modifications implies an active mechanism of histone replacement [48]. Indeed, Nucleosome Assembly Protein-1 (NAP-1) is highly expressed in the nucleus of E11.5 PGCs[46].

4. Re-Establishment of Sex Specific DNA Methylation Patterns in PGCs

The erasure of DNA methylation by E13.5 represents the most basal methylation state during the ontogeny of the germ line [31]. It is hypothesized that this massive erasure of methylation specific marks is critical for the establishment of sex-specific, gametic methylation marks to facilitate embryo development in terminally differentiated gametes, as well as to prevent the inheritance of epimutations in gametes [49]. DNA methylation is reset in a sex-dependent manner and is mediated by the *de* novo DNA methyltransferase *Dnmt3a* and *Dnmt3L*, a Dnmt-related family member that lacks catalytic methylation activity [50] [51]. During this period of *de novo* methylation, parental imprints are re-established sex-specifically [50].

In males, DNA remethylation shortly begins after genome wide cytosine erasure. From E14.5-birth, PGCs undergo quiescence, followed by re-methylation. Upon birth, these germ cells exit quiescence, and begin to divide mitotically, giving rise to the founding spermatogonial stem cell population within the gonad. DNA methylation patterns are maintained once specified [52-54].

By contrast, female re-methylation occurs after birth. After genome-wide erasure, female PGCs immediately undergo cell cycle arrest in prophase I of meiosis. After birth, PGCs exist meiotic arrest, and female-specific methylation marks are deposited during pachytene prior to ovulation [52,55-57].

5. Mechanisms of DNA Demethylation

Multiple mechanisms can potentially reconcile the two disparate events in the changing DNA methylation landscape of PGCs. In mammals, loss of 5mC from the mammalian genome can occur via broad strategies in somatic cell types. First, 5mC can be lost gradually and indiscriminately from the genome through inactivation of the cell-cycle dependent maintenance methylation machinery that modifies 5mC patterns as cells divide. Maintenance DNA methylation is necessary to deposit the proper 5mC marks after DNA has been replicated and is dependent upon the DNA methyltransferase *Dnmt1.* DNA maintenance methylation is enzymatically mediated by the maintenance methyltransferase *Dnmt1*[26]. However, **replication-dependent loss** of 5mC can occur through inactivity or repression of *Dnmt1* or *Uhrf1*, a co-factor that tethers Dnmt1 to chromatin during S phase to maintain methylation marks [58].

In addition to the inactivation of maintenance methylation machinery, 5mC can be removed from DNA via an active process involving chemical modification of the base to mediate active excision and replacement with unmodified cytidine. Indeed, mounting evidence from lower organisms has indicated a requirement for DNA repair mechanisms in mediating active demethylation, including base excision repair (BER) [47,59]. Chemical modification of 5mC can occur via deamination or hydroxylation.

Deamination of 5mC

Deamination of 5mC is achieved by the *Activation induced cytidine deaminase (Aicda) / Apobec* family of enzymes, and can results in the generation of a thymine base, resulting in a T-G mismatch in the DNA *in vitro* [60]. Removal of this base substitution could be initiated and resolved by specific glycosylases of the BER pathway, namely *Mbd4* and ***Thymine DNA Glycosylase (Tdg)*** [61,62]. Demethylation through deamination is a demonstrated phenomenon in zebrafish embryos [62]. In mammalian PGCs, *Aicda* is thought to play at least a minor role in genome-wide demethylation, with Aicda-/- PGCs having slightly higher methylation levels than wild type PGCs at E13.5 [31].

Hydroxylation of 5mC

5mC can also be modified by hydroxylation/oxidation. This conversion is mediated by the ***Ten-Eleven-Ten (Tet*)** dioxygenase family and results in the formation of **5-hydroxyethylcytosine (5hmC)**, a cytosine derivative that was first observed almost 60 years ago in bacteriophage [63]. Recent discovery of the Tet proteins have generated new insight into potential mechanisms for demethylation in multiple cell types that undergo dynamic epigenetic rearrangements. In mammals, 5hmC is generated by one of the three Tet enzymes (*Tet1, Tet2,* and *Tet3).* Tet proteins can also generate even further derivatives through subsequent hydroxylation conversion of 5hmC, resulting in the formation of 5-formylcytosine (5fC) or 5-carboxylcytosine (5caC) [64,65]. Though 5hmC-specific glycosylases have not been identified, the glycosylase *Tdg* has excision activity against 5caC [64,66]. However, replication dependent loss may also be coupled to 5hmC conversion by Tet proteins, as hydroxylation of 5hmC prevents association of Dnmt1 to hemi-hydroxymethylated DNA, resulting in eventual 5mC loss with division [67].

6. DNA Demethylation in Developmental Contexts

In addition to PGCs, DNA demethylation is implicated in the development of the early zygote, [5,68,69], early somatic progenitors [70], and brain [59]. DNA demethylation has also been linked to the establishment of tissue-specific gene expression patterns and hormone-induced demethylation to induced target gene expression in somatic cells [59,71].

DNA Demethylation in Zygotes and Pre-Implantation Embryos

The first demethylation event in early embryos occurs prior to zygote genome activation, and is mediated by hydroxylation of 5mC by Tet3 specifically on the paternal pronucleus. Active loss of paternal 5mC in early zygotes involves the activity of *Tet3* [72-74]. *Tet3* converts paternal 5mC bases to 5hmC, 5fC, and 5caC [64]. *Tet3* functions as a maternal effect gene, with *Tet3* null oocytes failing to initiate demethylation the paternal pronucleus and resulting in embryonic lethality [65,72-74]. Shielding of the activity of *Tet3* against promiscuous demethylation in the maternal pronucleus also requires maternal contribution of Stella by the oocyte to preserve maternal methylation and maintain methylation integrity of imprinted genes [75,76]. Furthermore, Stella's role to preserve methylation of the maternal genome is at the chromatin level, where it binds to H3K9m2, which is highly enriched in maternal pronuclei and at imprinted loci in the paternal genome [76].

Interestingly, 5hmC conversion in paternal pronuclei does not appear to be exclusively linked to BER. Further hydroxylation of 5hmC to 5fC and 5caC results in passive loss of 5mC with cell division [68]. Nonetheless, evidence for an involvement of BER in resolving pronuclear demethylation has been provided by [47], who found chromatin-incorporated BER components

XRCC1 and Parp1 specifically in the paternal pronucleus and Step 2 PGCs. Furthermore, treatment of zygotes with BER small molecule inhibitors impeded demethylation of repetitive LINE elements, suggesting that locus-specific removal of methylated cytosines from paternal pronucei is also possible. Despite its importance in pre-implantation development, *Tet3* is not involved in epigenetic reprogramming in PGCs because it is not expressed, but Tet1 and Tet2 were expressed in both Step1 and Step 2 PGCs [47,72].

After active paternal demethylation, the gametic pronuclei fuse resulting in embryonic genome activation. Genome-wide quantitation of methylation via reduced representation bisulfite sequencing (RRBS) provided insight into methylation regulaton with cell cleavage of early embryos [36]. First, profiling of sperm revealed that they are hypermethylated relative to oocytes. Second, after paternal demethylation by *Tet3*, DNA methylation in zygotes and early cleavage states is characterized by subtle progressive decreases in methylation levels, consistent with replication-dependent loss via cell division [36]. Furthermore, loss of methylation generally comes from cytosines that were initially methylated in sperm, and the methylation state of maternally derived cytosines remained relatively unchanged. This progressive loss in methylation continues until the ICM stage, where basal methylation is at its lowest in embryos [36]. Uncoupling the role of 5hmC in active removal as well as replication-dependent loss is currently unknown, but combined these data suggest that both processes are at play prior to blastocyst establishment [36,47,68,72,74].

Tet1 and Tet2

Tet1 and *Tet2* as well as 5hmC become highly expressed the inner cell mass of blastocysts and embryonic stem cells (ESCs) and decreases with differentiation [77]. Curiously, the single

knockout mice of either *Tet1* or *Tet2* develop normally, with mutants being runted and sub-fertile [77-79]. In ESCs *Tet1* and *Tet2* are directly regulated by binding of the master pluripotency regulators Oct4 and Sox2 in their promoter regions [77]. Interestingly, while Tet1-depleted ESCs are still capable of tri-lineage potential, differentiation is skewed to trophectoderm and mesendoderm lineages, suggesting that Tets and by extension 5hmC may regulate fate choice through demethylation mechanisms [65,77]. Indeed, loss of *Tet1* is associated with hypermethylation of both pluripotency and somatic lineage-associated genes [65,77]. Tet1 also binds CGIs regardless of their methylation status [80]. Tet1 ChIP studies have found that Tet1 is a CGI binding protein, with enriched occupancy at regions that tend to be unmethylated or lowly methylated, where it has been proposed to antagonize *de novo* methylation and the accumulation of repressive histone modifications [80,81]. In E13.5 female PGCs, Tet1 regulates locus-specific methylation of meiotic genes, and its deletion results in decreased demethylation of meiotic genes and unresolved DNA breaks during meiosis [79]. Loss of *Tet1* alone is insufficient to drastically affect genome wide demethylation by E13.5, leaving the possibility that functional redundancy between Tet1 and Tet2 may exist. This is reinforced by Koh et al., who found that overall levels of 5hmC in ESCs is unaffected by single knockdown of either Tet1 or Tet2, but are reduced upon depletion of both [77].

Thymine DNA Glycosylase

Removal of Tet-modified 5mC is regulated by BER through *Tdg*. Tdg, in addition to being the primary T-G mismatch enzyme, can also recognize 5caC and 5hmU-modified 5mC in vitro and in vivo [61,64]. Unlike the knockouts of any of the other DNA glycosylases [82,83], including *Mbd4, Smug1*, all of which are fertile and have mild-to-none developmental phenotypes, *Tdg* deletion results in embryonic lethality at E11.5 and is associated with multiple defects in histone

modifications and DNA methylation patterns in early development [61,84]. In mouse embryonic fibroblasts (MEFs), Tdg regulates p300-mediated transcriptional activation in addition to DNA methylation patterns. ChIP and BS-PCR experiments showed that Tdg is associated with developmentally associated genes, and that the loss of Tdg correlates with gene repression and hypermethylation of CGIs at these loci [84]. Importantly, another study also linked Tdg to regulation of imprinted genes in PGCs. BS-PCR of the differentially methylated region of *Igf2* revealed hypermethylation in Reprogramming I-stage PGCs, leading to the hypothesis that Tdg functions to maintain allele-specific methylation of imprinted genes [61].

From these studies, 5hmC and Tet proteins play a clear role in the maintenance of epigenetic plasticity in ESCs to maintain the capacity for self-renewal and pluripotency. And while Tet3 has a clear role in genome-wide modification of the paternal pronucleus in early zygotic development, genome-wide studies of cells actively undergoing massive demethylation have been limited [36,70] and none have identified a primarily Tet-driven mechanism to mediate genome-wide demethylation.

7. Evidence for Tet-Mediated Demethylation in PGCs

Mounting evidence indicates that an active demethylation occurs in Reprogramming I-stage PGCs. First, the loss of 5mC signal from PGC nuclei at E8.0 strikingly occurs in the majority of PGCs within 24 hours, but has only been characterized by immunofluorescence [29]. Second, tracing of 5mC signal demonstrates that the initial decreases in 5mC appear to be maintained until Reprogramming II, when a further drop in immunofluorescent signal is observed, suggesting a coordinated and regulated process that may not solely involve replication-dependent loss of 5mC [29]. The partial rather than progressive loss of 5mC may indicate bursts

of enzymatic conversion or active removal of 5mC at specific phases of PGC development [6,29]. However, given recent studies it is possible that Tet catalysis of 5hmC may be mutually exclusive from a requirement for BER [68]. Nonetheless, the accumulation of PGCs in G2 phase of the cell cycle at E8.25--shortly after their exit from the allantois, and the onset of demethylation—may serve to facilitate global reduction of 5mC before the cells begin to divide as they approach the gonad [45].

5hmC as a potential intermediate on the path to demethylation in PGCs is a tantalizing hypothesis. First, 5hmC generation is dependent upon pre-existing 5mC, with no evidence of 5hmC occurring *de novo* [85]. Second, 5hmC cannot be used as a template for Dnmt1 to maintain 5mC patterns. Even if BER is not involved in Step 1 demethylation, 5hmC can still induce demethylation in PGCs via potentiating replication dependent loss of 5mC. Third, although no DNA glycosylase with 5hmC excision activity has been identified, *Tdg* can excise 5caC *in vitro* and *in vivo* [64,66,86]. Finally, 5hmC can be subjected to deamination, generating 5hmU, another substrate that can be excised by *Tdg* [59]. Therefore, regardless of the final mechanism of modified 5mC removal, oxidation of 5mC may be a prerequisite for all subsequent demethylation pathways.

8. Generating PGCs *in vitro* from Embryonic Stem Cells

Challenges to understanding germ line biology during Step 1 reprogramming has been hampered by a number of technical obstacles that have been difficult to overcome. First, the number of PGCs that exist in vivo during Step 1 range from 40-200 cells per embryo between E7.5-E10.5 (Figure 1-4), in contrast to the thousands of PGCs that can be isolated at Step 2-stages of development. Second, PGCs during this phase do not localize to a discreet organ,

making isolation of the population difficult. Finally, there are few reporter alleles that facilitate specific isolation and purification of PGC populations by cell sorting. Therefore, the use of ESC differentiation to generate **PGCs *in vitro* (iPGCs)** has become an attractive model to study early PGC development [87-90].

In recent years multiple labs have generated PGC-like cells from ESCs, pluripotent cell lines derived from the inner cell mass of mouse and human blastocysts [87,91]. However, modeling of the progressive developmental steps that are undertaken by germ cells during their development have not been well studied, with variable results from multiple labs. Wei and colleagues have demonstrated that different ESC differentiation methods can give rise to PGCs with distinct gene expression patterns, which may represent subpopulations of PGCs with distinct developmental progression statuses [92]. Other *in vitro* systems have been used to screen for regulators of PGC development, and have provided insight into mechanisms of *Blimp1* induction [93], as well as the role of other PGC-associated genes [88,94-96]. The establishment of a model that capable of faithfully capturing in vitro derived Step 1-stage PGCs at a single cell resolution has not been reported.

Assessment of the epigenetic status of PGCs derived from ESCs as shown that they can undergo partial but incomplete demethylation at imprinted genes, suggesting that most ESC differentiation systems can only give rise to immature and pre-gonadal PGCs [88,90,97-99]. These studies have also shown that further demethylation of imprinted genes cannot be achieved through prolonged differentiation, indicating that the differentiation microenvironment *in vitro* is insufficient to support further the subsequent epigenetic reprogramming events that occur in

gonadal-stage PGCs [88,93]. Indeed, the requirement of a somatic, gonadal *in vivo* niche was necessary to generate functional sperm [98] and oocytes [97] from ESCs.

PGC Isolation Methods from ESCs

To isolate presumptive PGCs, multiple markers have been used to delineate in vitro PGCs in non-directed differentiation strategies, including transgenic reporters for multiple PGC-associated genes (Table 1-1). Other studies, including the one presented in Chapter 2, have devised isolation strategies that rely on unique expression patterns of cell surface markers on the surface of in vitro PGCs, and can therefore facilitate study of PGC development using pre-existing mutant ESC lines. The establishment of robust PGC differentiation systems requires a stringent test of PGC identity, which has been lacking from many of the studies claiming to generate PGCs in vitro. Furthermore, the characterization of mechanisms of PGC development using in vitro models have been limited, and is likely attributable to lack of a model that can faithfully capture stage-specific PGCs.

Available Assays for Germ Cell Identity of ES-Derived Cells

Due to the overlap in expression of many ES-associated genes in the germ line, the identification of pre-gonadal embryonic PGCs has been difficult to establish by many labs. The most widely used assay of PGC identity *in vitro* takes advantage of the differential response ESCs and PGCs have to retinoic acid (RA), a morphogen that promotes differentiation of ESCs and self-renewal of PGCs in culture [87,100]. This assay has been used to establish PGC identity by multiple laboratories in their differentiation strategies, but does not preclude the possibility that the resultant differentiations have not yielded failed-to-differentiate ESCs that appear to be

PGC-like [87,93,101]. To ensure that putative isolated cells are putative PGCs, regardless of the isolation strategy, necessitates a more stringent assay of PGC identity.

Additional assays of germ cell identity test the functional capacity of gametes derived in vitro. Current assays to test the functional differentiation capacity of gametes involves transplantation of in vitro derived germ cells into the lumen of the seminiferous tubules in testes of genetically sterile newborn mice. In this assay, endogenous PGCs as young as E8.5 can engraft in the somatic gonad, giving rise to robust foci of spermatogenesis after 6-8 weeks [102]. Epiblast cells induced to PGC fate through chemically defined media are also able to generate live, fertile young [103]. To date, only one laboratory has been successfully demonstrated that ESCs can give rise to transplantable PGCs that can generate viable gametes, both male and female. These strategies required directed differentiation through an epiblast intermediate via chemically defined factors and recombinant proteins [97,98].

Aims and Significance of the Dissertation

Taken together, the studies discussed above indicate that early events in PGC life are largely not understood. Study of mediators of PGC specification and early DNA demethylation necessitates the generation of a thoroughly vetted model that exclusively has features of Step 1-PGCs. The importance of understanding these mechanisms are crucial to identifying the intrinsic pathways that underlie genomic totipotency. Furthermore, detailed understanding of these mechanisms may be useful in applying these processes to artificially inducing epigenetic plasticity during cellular reprogramming of somatic cells to induced pluripotent stem cells.

Therefore, the aims of this thesis were to:

1. Devise an ESC-based system to generate Reprogramming I stage-specific, pre-gonadal PGCs *in vitro* (iPGCs) without the use of integrated reporter transgenes.

2. Devise new means to interrogate PGC identity and staging via functional assays and single cell gene expression technology.

3. Characterize mechanisms of Step 1 epigenetic reprogramming, in particular genome-wide DNA demethylation, and to characterize potentiators of this epigenetic remodeling using an iPGC differentiation model system.

Taken together, these studies will demonstrate that the development of a novel differentiation strategy is capable of isolating PGCs in vitro, with transcriptional and epigenetic features specifically reminiscent of a defined window in PGC development. Development of new assays to evaluate PGC identity and maturation were employed, including genetic functionality tests and analysis of single cell gene expression. Finally, I use this system to functionally address candidate enzymes in the genome-wide DNA demethylation process.

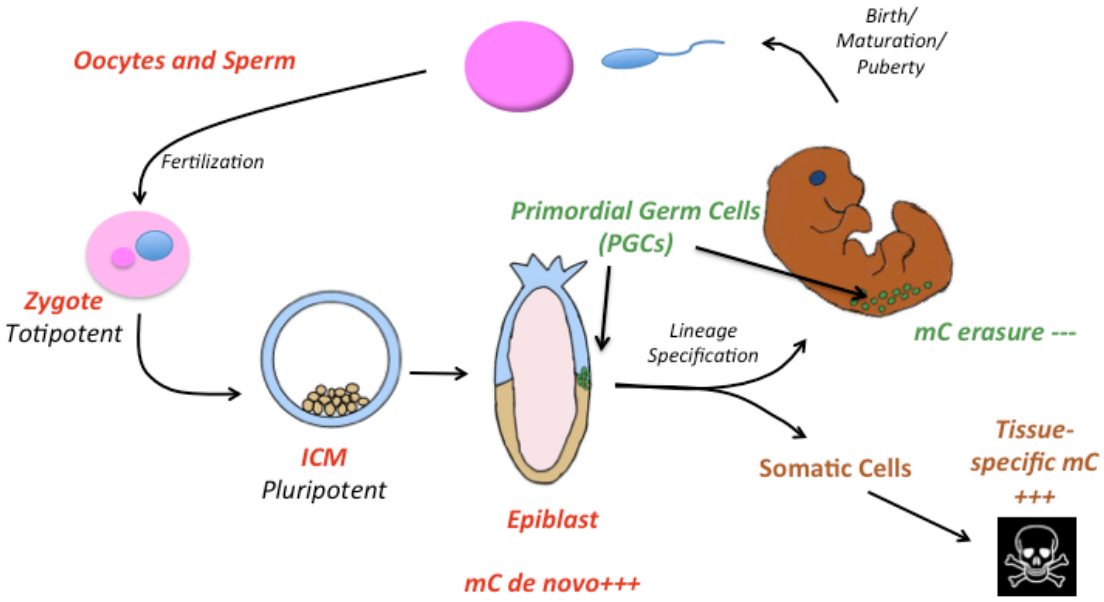

Figure 1-1. Overview of Germ Line Development.

Mature germ cells, termed sperm and oocyte, can generate new offspring via fertilization and the establishment of the totipotent zygote. Subsequent development of the zygote leads to the establishment of the blastocyst, containing the inner cell mass (ICM) which will give rise to all tissues of the embryo proper. During epiblast development, massive genome-wide addition of methyl groups to cytosine bases in the DNA takes place and is thought to facilitate lineage specification. The epiblast will give rise to both somatic cells and PGCs. Somatic cells retain this 5mC deposition and commit to their respective lineages, and serve their function until they invariably die. In contrast, while PGCs are initially specified from methylated epiblast cells, they undergo subsequent global depletion of 5mC prior to birth. Upon birth and puberty, germ line cells will undergo sex-specific development to create new haploid gametes to generate new offspring. Germ cells that successfully accomplish fertilization will not die but instead contribute to the development of the next generation.

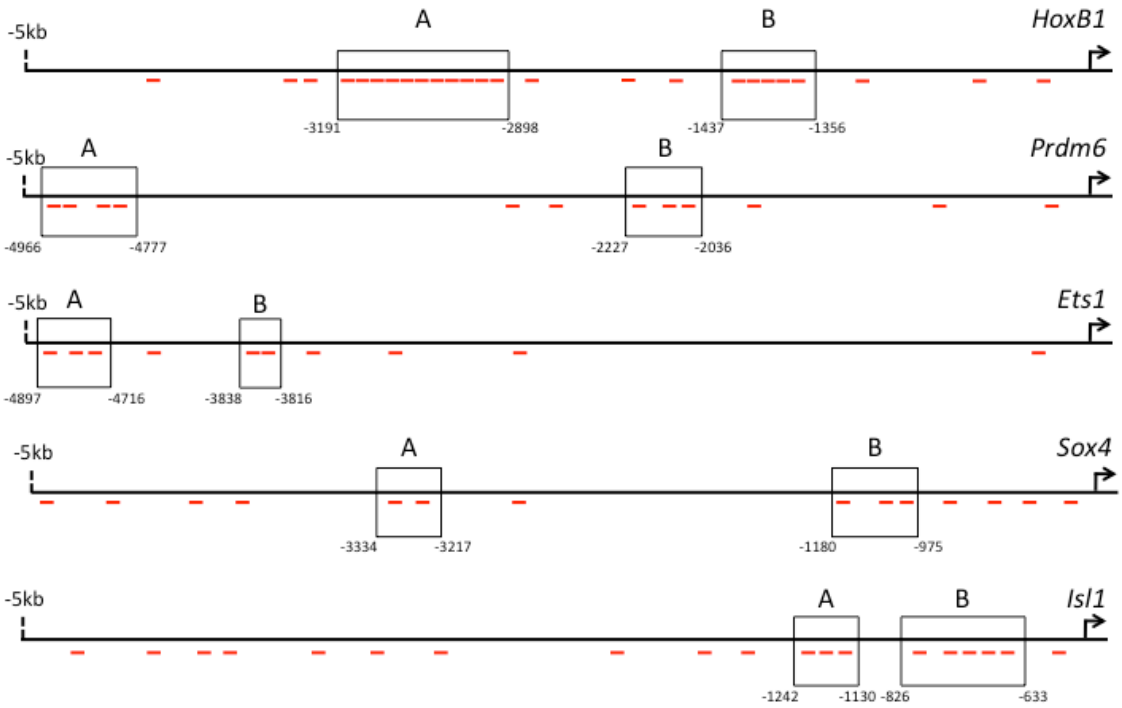

Figure 1-2. Analysis of GAAAG motifs in genes regulated by Blimp1 identified by Kurimoto and colleagues.

Shown are promoter schematics of 5kb upstream of the transcriptional start site. Boxes marked A and B refer to dense clusters of GAAAG motifs observed in close proximity to each other in the promoters indicated.

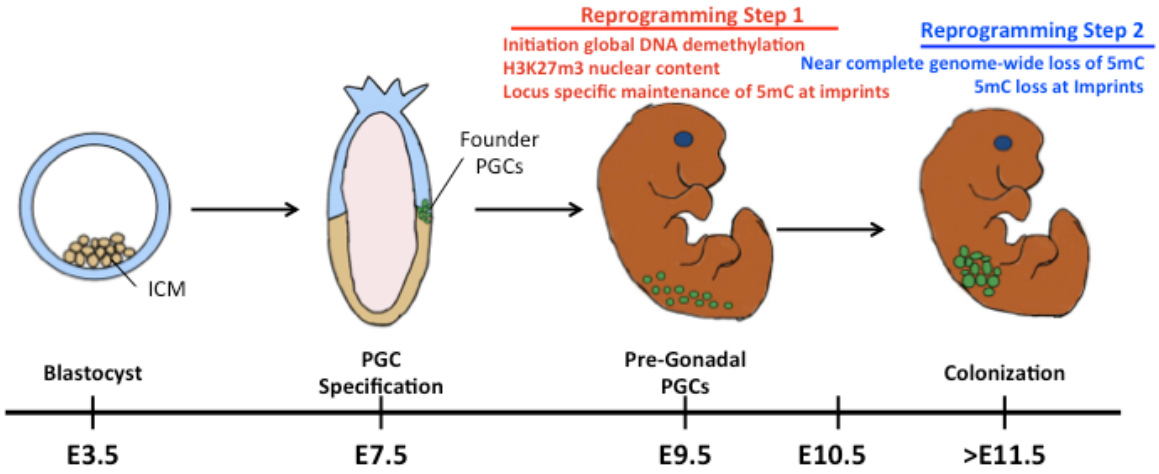

Figure 1-3. Epigenetic Remodeling in Early PGC Development.

Step 1 reprogramming begins at E8.0 and extends until E10.5. DNA demethylation is observed to be globally reduced by immunofluorescence, but maintenance of methylation at imprinted genes has been observed via PCR amplification of bisulfite-treated DNA from PGCs (Hajkova 2002). Step 2 reprogramming occurs from after E10.5-13.5 and is characterized by erasure of imprints.

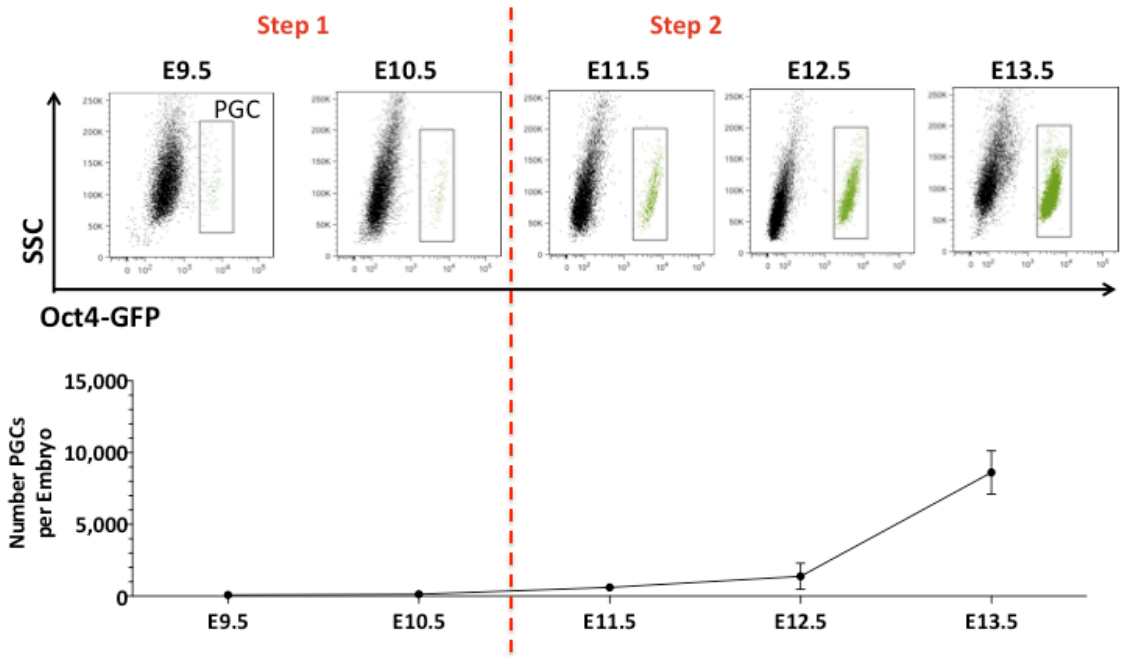

Figure 1-4. Flow cytometry analysis of Oct4-GFP embryos from E9.5-E13.5.

PGCs are isolated by GFP expression (Box). PGCs increase in number with development. Numbers of PGCs available from Step 1 are prohibitive, resulting in ~100 cells/embryos at E9.5 and ~200 at E10.5.

Year	First Author	Diff Method	Isolation Strategy	Epigenetic Charac?	PGC Staging?	Identity Assay?
2003	Hubner	Adherent	*Oct4.tPE-gfp*	None	None	None
2003	Toyooka	EBs, EB+BMP4	*Myh-gfp/LacZ*	None	None	Myh+ cells aggregated with E12.5-13.5 gonads and transplanted give ride to ES-derived sperm
2004	Geijsen	EB 9day	*SSEA1*	*Igf2r* Restriction on EGCs	None	Injection of haploid EB-derived PGC develop to 2cell stage
2006	Payer	Hanging Drop EB then Shaking EB	*Stella-gfp*	None	None	None
2006	Mizuno	EB	*Myh-LacZ*	None	None	None
2007	Qing	EB and coculture w/ newborn granulosa cells	*SSEA1, cKit*	None	None	None
2008	Wei	Adherent and EB	*Stella-gfp*	BS-PCR	None	None
2009	Haston	EB 35day	*Oct4.tPE-gfp*	BS-PCR	None	None
2009	West	EB 7day	*Stella-gfp*	BS-PCR	E10 PGC (MA)	KD PGC determinants
2009	Nicholas	EB 15day	*Dazl-gfp*	None	None	None
2009	Young	EB 10day	*Oct4.tPE-gfp*	BS-PCR on EGCs	None	None
2009	Yu	Adherent	Colony picking	None	None	"Sperm-like" cells fertilize oocytes: 2 cell stage
2010	Imamura	EB, Adherent, Aggregate w/ BMP+ cells	Oct4-gfp/Myh-rfp	None	None	RA Assay
2011	Tedesco	Adherent	SSEA1, AP	None	None	Defective Assembly of Meiotic Proteins
2011	**Hayashi**	**CDM**	**Blimp/Stella & SSEA1/Itgb3**	**BS-PCR, 5mC Loss, H3K27m3+**	**E9.5 (MA)**	**Neonatal transplantation generates spermatogenic foci**
2011	**Vincent**	**Hanging Drop EB 6day**	**SSEA1&Oct4/cKit^{bright}**	**BS-PCR, H3K27m3+**	**E9.5-E10.5 (SCP)**	**Blimp1 KO, RA Assay**
2012	**Hayashi**	**CDM**	**Blimp/Stella & SSEA1/Itgb3**	**BS-PCR, 5mC Loss, H3K27m3+**	**"**	**Transplantation gives ride to oocytes capable of producing fertile young**

Table 1-1. Summary of Pluripotent Stem Cell-Based Studies to Generate iPGCs.
Multiple groups have reported the generation of PGCs from ESCs with varying degrees of stringency, ranging from expression of reporter molecules, transgenes, rudimentary characterization of developmental staging and epigenetic progression. Notably, many groups have failed to assay PGC identity. Studies highlighted in blue, including the one in Chapter 2, have addressed all of these aspects of PGC characterization in their differentiation systems. EB=embryoid body, CDM=chemically defined media, MA=microarray.

REFERENCES

1 Hayashi, K., de Sousa Lopes, S. M. & Surani, M. A. Germ cell specification in mice. *Science* **316**, 394-396, doi:10.1126/science.1137545 (2007).

2 Saitou, M., Barton, S. C. & Surani, M. A. A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300, doi:10.1038/nature00927 (2002).

3 Tanaka, S. S. & Matsui, Y. Developmentally regulated expression of mil-1 and mil-2, mouse interferon-induced transmembrane protein like genes, during formation and differentiation of primordial germ cells. *Mech Dev* **119 Suppl 1**, S261-267 (2002).

4 Ohinata, Y. *et al.* Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213, doi:10.1038/nature03813 (2005).

5 Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature reviews. Genetics* **2**, 21-32, doi:10.1038/35047554 (2001).

6 Hackett, J. A., Zylicz, J. J. & Surani, M. A. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends in genetics : TIG* **28**, 164-174, doi:10.1016/j.tig.2012.01.005 (2012).

7 Arnold, S. J., Maretto, S., Islam, A., Bikoff, E. K. & Robertson, E. J. Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. *Dev Biol* **296**, 104-118, doi:10.1016/j.ydbio.2006.04.442 (2006).

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

9 Hayashi, K. *et al.* SMAD1 signaling is critical for initial commitment of germ cell lineage from mouse epiblast. *Mech Dev* **118**, 99-109 (2002).

10 Lawson, K. A. *et al.* Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13**, 424-436 (1999).

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

12 Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Molecular endocrinology (Baltimore, Md.)* **14**, 1053-1063 (2000).

13 Ying, Y., Qi, X. & Zhao, G. Q. Induction of primordial germ cells from murine epiblasts by synergistic action of BMP4 and BMP8B signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7858-7862, doi:10.1073/pnas.151242798 (2001).

14 Keller, A. D. & Maniatis, T. Identification and characterization of a novel repressor of beta-interferon gene expression. *Genes Dev* **5**, 868-879 (1991).

15 John, S. A., Clements, J. L., Russell, L. M. & Garrett-Sinha, L. A. Ets-1 regulates plasma cell differentiation by interfering with the activity of the transcription factor Blimp-1. *J Biol Chem* **283**, 951-962, doi:10.1074/jbc.M705262200 (2008).

16 Kuo, T. C. & Calame, K. L. B lymphocyte-induced maturation protein (Blimp)-1, IFN regulatory factor (IRF)-1, and IRF-2 can bind to the same regulatory sites. *J Immunol* **173**, 5556-5563 (2004).

17 Piskurich, J. F. *et al.* BLIMP-1 mediates extinction of major histocompatibility class II transactivator expression in plasma cells. *Nat Immunol* **1**, 526, doi:doi:10.1038/82788 (2000).

18 Shaffer, A. L. *et al.* Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17**, 51-62 (2002).

19 Shapiro-Shelef, M. *et al.* Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* **19**, 607-620 (2003).

20 Robertson, E. J. *et al.* Blimp1 regulates development of the posterior forelimb, caudal pharyngeal arches, heart and sensory vibrissae in mice. *Development* **134**, 4335-4345, doi:10.1242/dev.012047 (2007).

21 Vincent, S. D. *et al.* The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* **132**, 1315-1325, doi:10.1242/dev.01711 (2005).

22 Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* **22**, 1617-1635, doi:10.1101/gad.1649908 (2008).

23 Doody, G. M. *et al.* An extended set of PRDM1/BLIMP1 target genes links binding motif type to dynamic repression. *Nucleic Acids Research* **38**, 5336-5350, doi:10.1093/nar/gkq268 (2010).

24 Ancelin, K. *et al.* Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat Cell Biol* **8**, 623-630, doi:10.1038/ncb1413 (2006).

25 Tee, W. W. *et al.* Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes Dev* **24**, 2772-2777, doi:10.1101/gad.606110 (2010).

26 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

27 Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371-382 (1987).

28 Monk, M. Methylation and the X chromosome. *BioEssays* **4**, 204-208, doi:10.1002/bies.950040505 (1986).

29 Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**, 440-458, doi:10.1016/j.ydbio.2004.11.025 (2005).

30 Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23 (2002).

31 Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101-1105, doi:10.1038/nature08829 (2010).

32 Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular cell* (2012).

33 Lei, H. *et al.* De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195-3205 (1996).

34 Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139**, 15-31, doi:10.1242/dev.050849 (2012).

35 Guibert, S., Forne, T. & Weber, M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome research* **22**, 633-641, doi:10.1101/gr.130997.111 (2012).

36 Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-344, doi:10.1038/nature10960 (2012).

37 Richardson, B. E. & Lehmann, R. Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat Rev Mol Cell Biol* **11**, 37-49, doi:10.1038/nrm2815 (2010).

38 Orr-Urtreger, A. *et al.* Developmental expression of c-kit, a proto-oncogene encoded by the W locus. *Development* **109**, 911-923 (1990).

39 Gu, Y., Runyan, C., Shoemaker, A., Surani, A. & Wylie, C. Steel factor controls primordial germ cell survival and motility from the time of their specification in the allantois, and provides a continuous niche throughout their migration. *Development* **136**, 1295-1303, doi:10.1242/dev.030619 (2009).

40 Runyan, C. *et al.* Steel factor controls midline cell death of primordial germ cells and is essential for their normal proliferation and migration. *Development* **133**, 4861-4869, doi:10.1242/dev.02688 (2006).

41 Godin, I. *et al.* Effects of the steel gene product on mouse primordial germ cells in culture. *Nature* **352**, 807-809, doi:10.1038/352807a0 (1991).

42 Boldajipour, B. *et al.* Control of chemokine-guided cell migration by ligand sequestration. *Cell* **132**, 463-473, doi:10.1016/j.cell.2007.12.034 (2008).

43 Ara, T. *et al.* Impaired colonization of the gonads by primordial germ cells in mice lacking a chemokine, stromal cell-derived factor-1 (SDF-1). *Proceedings of the National*

Academy of Sciences of the United States of America **100**, 5319-5323, doi:10.1073/pnas.0730719100 (2003).

44 Molyneaux, K. A. *et al.* The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. *Development* **130**, 4279-4286 (2003).

45 Seki, Y. *et al.* Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134**, 2627-2638, doi:10.1242/dev.005611 (2007).

46 Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-881, doi:10.1038/nature06714 (2008).

47 Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78-82, doi:10.1126/science.1187945 (2010).

48 Tam, P. P. & Snow, M. H. Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. *J Embryol Exp Morphol* **64**, 133-147 (1981).

49 Aravin, A. A. & Bourc'his, D. Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev* **22**, 970-975, doi:10.1101/gad.1669408 (2008).

50 Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903, doi:10.1038/nature02633 (2004).

51 Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536-2539, doi:10.1126/science.1065848 (2001).

52 Chaillet, J. R., Vogt, T. F., Beier, D. R. & Leder, P. Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. *Cell* **66**, 77-83 (1991).

53 Kato, Y. *et al.* Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* **16**, 2272-2280, doi:10.1093/hmg/ddm179 (2007).

54 Lees-Murdock, D. J., De Felici, M. & Walsh, C. P. Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics* **82**, 230-237 (2003).

55 Kono, T., Obata, Y., Yoshimzu, T., Nakahara, T. & Carroll, J. Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. *Nat Genet* **13**, 91-94, doi:10.1038/ng0596-91 (1996).

56 Lucifero, D., Mann, M. R., Bartolomei, M. S. & Trasler, J. M. Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum Mol Genet* **13**, 839-849, doi:10.1093/hmg/ddh104 (2004).

57 Lucifero, D., Mertineit, C., Clarke, H. J., Bestor, T. H. & Trasler, J. M. Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* **79**, 530-538, doi:10.1006/geno.2002.6732 (2002).

58 Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764, doi:10.1126/science.1147939 (2007).

59 Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423-434, doi:10.1016/j.cell.2011.03.022 (2011).

60 Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* **279**, 52353-52360, doi:10.1074/jbc.M407695200 (2004).

61 Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67-79, doi:10.1016/j.cell.2011.06.020 (2011).

62 Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201-1212, doi:10.1016/j.cell.2008.11.042 (2008).

63 Wyatt, G. R. & Cohen, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *The Biochemical journal* **55**, 774-782 (1953).

64 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).

65 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133, doi:10.1038/nature09303 (2010).

66 Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature chemical biology* **8**, 328-330, doi:10.1038/nchembio.914 (2012).

67 Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* **40**, 4841-4849, doi:10.1093/nar/gks155 (2012).

68 Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell research* **21**, 1670-1676, doi:10.1038/cr.2011.189 (2011).

69 Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**, 475-478 (2000).

70 Shearstone, J. R. *et al.* Global DNA demethylation during mouse erythropoiesis in vivo. *Science* **334**, 799-802, doi:10.1126/science.1207306 (2011).

71 Ma, D. K. *et al.* Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* **323**, 1074-1077, doi:10.1126/science.1166859 (2009).

72 Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610, doi:10.1038/nature10443 (2011).

73 Iqbal, K., Jin, S. G., Pfeifer, G. P. & Szabo, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3642-3647, doi:10.1073/pnas.1014033108 (2011).

74 Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature communications* **2**, 241, doi:10.1038/ncomms1240 (2011).

75 Nakamura, T. *et al.* PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat Cell Biol* **9**, 64-71, doi:10.1038/ncb1519 (2007).

76 Nakamura, T. *et al.* PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**, 415-419, doi:10.1038/nature11093 (2012).

77 Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell* **8**, 200-213, doi:10.1016/j.stem.2011.01.008 (2011).

78 Dawlaty, M. M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell stem cell* **9**, 166-175, doi:10.1016/j.stem.2011.07.010 (2011).

79 Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature*, doi:10.1038/nature11709 (2012).

80 Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular cell* **42**, 451-464, doi:10.1016/j.molcel.2011.04.005 (2011).

81 Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389-393, doi:10.1038/nature09934 (2011).

82 Kemmerich, K., Dingler, F. A., Rada, C. & Neuberger, M. S. Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice. *Nucleic Acids Res* **40**, 6016-6025, doi:10.1093/nar/gks259 (2012).

83 Millar, C. B. *et al.* Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297**, 403-405, doi:10.1126/science.1073354 (2002).

84 Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419-423, doi:10.1038/nature09672 (2011).

85 Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398-402, doi:10.1038/nature10008 (2011).

86 Hashimoto, H., Hong, S., Bhagwat, A. S., Zhang, X. & Cheng, X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res* **40**, 10203-10214, doi:10.1093/nar/gks845 (2012).

87 Geijsen, N. *et al.* Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* **427**, 148-154, doi:10.1038/nature02247 (2004).

88 Haston, K. M., Tung, J. Y. & Reijo Pera, R. A. Dazl functions in maintenance of pluripotency and genetic and epigenetic programs of differentiation in mouse primordial germ cells in vivo and in vitro. *PLoS ONE* **4**, e5654, doi:10.1371/journal.pone.0005654 (2009).

89 Sabour, D. *et al.* Identification of genes specific to mouse primordial germ cells through dynamic global gene expression. *Hum Mol Genet*, doi:10.1093/hmg/ddq450 (2010).

90 Vincent, J. J. *et al.* Single cell analysis facilitates staging of Blimp1-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS One* **6**, e28960, doi:10.1371/journal.pone.0028960 (2011).

91 Clark, A. T. *et al.* Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. *Hum Mol Genet* **13**, 727-739, doi:10.1093/hmg/ddh088 (2004).

92 Wei, W. *et al.* Primordial germ cell specification from embryonic stem cells. *PLoS ONE* **3**, e4013, doi:10.1371/journal.pone.0004013 (2008).

93 West, J. A. *et al.* A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* **460**, 909-913, doi:10.1038/nature08210 (2009).

94 Nicholas, C. R. *et al.* Characterization of a Dazl-GFP germ cell-specific reporter. *Genesis* **47**, 74-84, doi:10.1002/dvg.20460 (2009).

95 Toyooka, Y. *et al.* Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mechanisms of Development* **93**, 139-149 (2000).

96 Yu, Z. *et al.* Dazl promotes germ cell differentiation from embryonic stem cells. *J Mol Cell Biol* **1**, 93-103, doi:10.1093/jmcb/mjp026 (2009).

97 Hayashi, K. *et al.* Offspring from Oocytes Derived from in Vitro Primordial Germ Cell-Like Cells in Mice. *Science*, doi:10.1126/science.1226889 (2012).

98 Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532, doi:10.1016/j.cell.2011.06.052 (2011).

99 West, J. A., Park, I.-H., Daley, G. Q. & Geijsen, N. In vitro generation of germ cells from murine embryonic stem cells. *Nat Protoc* **1**, 2026-2036, doi:10.1038/nprot.2006.303 (2006).

100 Koshimizu, U., Watanabe, M. & Nakatsuji, N. Retinoic acid is a potent growth activator of mouse primordial germ cells in vitro. *Developmental Biology* **168**, 683-685, doi:10.1006/dbio.1995.1113 (1995).

101 Imamura, M. *et al.* Induction of primordial germ cells from mouse induced pluripotent stem cells derived from adult hepatocytes. *Mol. Reprod. Dev.* **77**, 802-811, doi:10.1002/mrd.21223 (2010).

102 Chuma, S. *et al.* Spermatogenesis from epiblast and primordial germ cells following transplantation into postnatal mouse testis. *Development* **132**, 117-122, doi:10.1242/dev.01555 (2005).

103 Ohinata, Y. *et al.* A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. *Cell* **137**, 571-584, doi:10.1016/j.cell.2009.03.014 (2009).

CHAPTER 2

SINGLE CELL ANALYSIS FACILITATES STAGING OF *BLIMP1*-DEPENDENT PRIMORDIAL GERM CELLS DERIVED FROM MOUSE EMBRYONIC STEM CELLS

Single Cell Analysis Facilitates Staging of *Blimp1*-Dependent Primordial Germ Cells Derived from Mouse Embryonic Stem Cells

John J. Vincent[1-3], Ziwei Li[1-2], Serena A. Lee[1-2], Xian Liu[1], Marisabel O. Etter[1-2], Silvia V. Diaz-Perez[1-2], Sara K. Taylor[1], Sofia Gkountela[1-2], Anne G. Lindgren[1-2], Amander T. Clark[1-4]

1 Department of Molecular Cell and Developmental Biology, 2 Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, 3 Molecular Biology Institute, 4 Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA, 90095; USA

ABSTRACT

The cell intrinsic programming that regulates mammalian primordial germ cell (PGC) development in the pre-gonadal stage is challenging to investigate. To overcome this we created a transgene-free method for generating PGCs *in vitro* (iPGCs) from mouse embryonic stem cells (ESCs). Using labeling for SSEA1 and cKit, two cell surface molecules used previously to isolate presumptive iPGCs, we show that not all SSEA1+/cKit+ double positive cells exhibit a PGC identity. Instead, we determined that selecting for cKit^{bright} cells within the SSEA1+ fraction significantly enriches for the putative iPGC population. Single cell analysis comparing SSEA1+/cKit^{bright} iPGCs to ESCs and embryonic PGCs demonstrates that 97% of single iPGCs co-express PGC signature genes *Blimp1, Stella, Dnd1, Prdm14* and *Dazl* at similar levels to e9.5-10.5 PGCs, whereas 90% of single mouse ESC do not co-express PGC signature genes. For the 10% of ESCs that co-express PGC signature genes, the levels are significantly lower than iPGCs. Microarray analysis shows that iPGCs are transcriptionally distinct from ESCs and repress gene ontology groups associated with mesoderm and heart development. At the level of chromatin, iPGCs contain 5-methyl cytosine bases in their DNA at imprinted and non-imprinted loci, and are enriched in histone H3 lysine 27 trimethylation, yet do not have detectable levels of Mvh protein, consistent with a Blimp1-positive pre-gonadal PGC identity. In order to determine whether iPGC formation is dependent upon Blimp1, we generated *Blimp1* null ESCs and found that loss of Blimp1 significantly depletes SSEA1/cKit^{bright} iPGCs. Taken together, the generation of *Blimp1*-positive iPGCs from ESCs constitutes a robust model for examining cell-intrinsic regulation of PGCs during the *Blimp1*-positive stage of development.

INTRODUCTION

The molecular events that regulate cell fate decisions in post-implantation mammalian embryonic development are largely uncharacterized due to the challenge in identifying and isolating small populations of specific precursor cells that are developmentally transient in the early embryo. In particular, precursors of the germ cell lineage are initially set aside as four to six cells in the murine embryonic epiblast, which proliferate and migrate through the primitive streak to generate the initial founder primordial germ cell (PGC) pool of approximately forty cells at the base of the allantois at embryonic (e) day e7.5 [1,2]. The PGCs migrate out of the allantois and into the embryonic hindgut endoderm at e8.0-8.5 where they continue to proliferate and begin to accumulate nuclear histone H3 lysine 27 trimethylation (H3K27m3) [3]. By e10.5-e11.0 a single embryo has approximately 1,000-2,000 PGCs, which exit the hindgut and begin colonization of the indifferent gonad and express Mvh protein [3-5].

The transcription factors that specify and sustain PGC identity prior to gonadal colonization are not well understood. One of the most characterized regulators of PGC fate is the transcriptional repressor *B-lymphocyte induced maturation protein 1 (Blimp1),* the transcriptional product of the *PRD1-BF1 and RIZ (PR) domain 1* (*Prdm1*) gene. Blimp1 expression is detected in epiblast-derived PGCs and persists until e11.5, when PGCs have colonized the gonad [2,6]. Loss of one *Prdm1* allele significantly reduces PGC numbers in the allantois, with the loss of both causing almost a complete loss of PGCs [2]. The major direct target of Blimp1 in PGCs is hypothesized to be *Hoxb1* [7]. However, direct binding of Blimp1 at the *Hoxb1* locus in PGCs has not been demonstrated.

The mechanism by which Blimp1 mediates gene repression is hypothesized to involve recruitment of the chromatin-remodeling enzyme Protein arginine methyltransferase 5 (Prmt5)

to chromatin [6]. However, genome-wide analysis of PGC chromatin is currently not feasible due to the challenge in performing chromatin immunoprecipitation on small cell numbers, necessitating the development of a scalable model to accurately capture the Blimp1-positive phase of PGC development.

The differentiation of pluripotent stem cells, including embryonic stem cells (ESCs), has emerged as a novel technology for generating sufficient numbers of embryonic progenitors at-scale to evaluate embryonic lineage development. A number of methods for identifying *in vitro* PGCs (iPGCs) have been described that mostly involve use of integrated fluorescent reporters, including *Oct4-delta-PE-Gfp* [8-12], *Stella-Gfp* [13-15], *Dazl-Gfp* [16] and *Mvh-LacZ/Rfp* transgenes [10,17,18]. A small number of studies have used Stage Specific Embryonic Antigen 1 (SSEA1) to enrich for germ cells [19,20], but the identity of PGCs from ESCs within the SSEA1+ fraction has not been interrogated at a single cell level. Furthermore, the majority of PGC differentiation studies have been designed to characterize the post-colonized Blimp1-negative PGC. Therefore, the goal of the current study is to generate a robust ESC differentiation system to acquire PGCs in the Blimp1-positive stage of development for future in-depth analysis of the pre-gonadal stage.

RESULTS

cKit^{bright} refines an Oct4+/SSEA1+ iPGC population in embryoid bodies

To identify pre-gonadal iPGCs with differentiation, we first used *Oct4-Gfp* ESCs [21] to generate hanging-drop embryoid bodies (EBs) containing 300 cells per drop (Figure 2-1A). EBs could be maintained for up to 8 days in this system (Figure 2-S1A), but cell viability decreased rapidly

after day 6 from 69% to 19% by day 8 (Figure 2-S1B). Using flow cytometry we show that Gfp is retained in the majority of cells in the first four days of differentiation (Figure 2-1B), reminiscent of sustained Oct4 expression *in vivo* in both PGCs and embryonic somatic cells up to e8.5 [22,23]. On day 5 of differentiation, we observed the emergence of a shoulder of Gfp^{bright} cells and the formation of a distinct Gfp+ peak by day 6 (Figure 2-1B, arrow).

To generate a transgene-free method of iPGC differentiation, we correlated expression of Oct4 protein in day 6 EBs derived from V6.5 ESCs with the cell surface marker SSEA1. In the embryo, SSEA1 is highly expressed on Blimp1-positive stage PGCs and PGC precursors derived from epiblast stem cells [24,25]. We found that Oct4 is co-expressed with SSEA1 in small cell clusters at day 6 of differentiation by immunofluorescence (Figure 2-1C). Given that Oct4 and SSEA1 are also expressed by undifferentiated ESCs, we used the membrane-localized tyrosine kinase receptor cKit to assist in further defining the iPGC population within either the SSEA1+ or Oct4+ fractions. *cKit* is highly expressed by endogenous PGCs from e7.25 to e13.5 [7,22,26] and is not expressed by epiblast cells [22]. Indeed, flow cytometry analysis of V6.5 ESC-derived EBs at day 6 of differentiation revealed a discreet a side population of SSEA1+/cKit+ cells (Figure 2-1D). A side population of Oct4-Gfp+/cKit+ cells was also identified beginning at day 6 of differentiation in EBs derived from *Oct4-gfp* ESCs, and this was sustained to day 8 (Figure 2-S1C).

To interrogate PGC identity in specific fractions of SSEA1+/cKit+ cells when the population is first identified at day 6, we used real time PCR to determine relative levels of PGC-expressed genes in discreet cKit+ fractions. These fractions include $SSEA1+/cKit^{bright}$ (green), $SSEA1+/cKit^{mid}$ (light blue), and $SSEA1+/cKit^{dim}$ (dark blue), and SSEA1-/cKit- cells (red) as a

negative control (Figure 2-1E). cKit^{bright} cells were selected based on increased signal intensity above the main population. PGC genes including *cKit*, *Blimp1*, *Stella* and *Dazl* were all enriched in the SSEA1+/cKit^{bright} fraction, with lower expression in the cKit^{mid} and cKit^{dim} fractions of SSEA1+ cells (Figure 2-1F). *Mvh* was also expressed in the SSEA1+/cKit^{bright} fraction, but was not specifically enriched in SSEA1+/cKit^{bright} cells relative to cKit^{mid} and cKit^{dim}. Furthermore, analysis of *Mvh* levels in SSEA1+/cKit^{bright} cells at day 8 of differentiation also did not show an increase relative to day 6 (data not shown). In contrast, transcription factors expressed in somatic cells such as *Hoxa1* and *Hoxb1* were highly expressed in SSEA1-/cKit- and SSEA1+/cKit^{dim} relative to SSEA1+/cKit^{bright} cells. Together, we conclude that not every SSEA1+/cKit+ cell in EBs at day 6 of differentiation is a putative PGC, and that selecting for cKit^{bight} cells enriches for the iPGCs beginning at day 6 of differentiation and persisting until day 8. We next determined if Oct4-gfp could further sub-fractionate the putative SSEA1+/cKit^{bright} putative iPGC population, and found equal enrichment of Oct4-gfp in all SSEA1+ cells regardless of cKit intensity (Figure 2-S1D). Therefore the use of the Oct4-Gfp reporter together with SSEA1 and cKit does not further refine the isolation of putative iPGCs, but instead shows that Oct4-Gfp and SSEA1 report the same population when used with cKit.

We next evaluated the yield of SSEA1+/cKit^{bright} or Oct4-gfp+/cKit^{bright} cells (called iPGCs) isolated at day 6 of differentiation (Figure 2-1G). We determined that 1-4% of total live EB cells exhibited an iPGC surface signature, and there was no statistical difference between genetic backgrounds (Figure 2-1G). Furthermore, gene expression profiling of Oct4-gfp+/cKit^{bright} cells from the *Oct4-gfp* line and SSEA1+/cKit^{bright} cells from J1 EBs at day 6 revealed enriched expression of *cKit, Stella, Blimp1, Dazl* and *Mvh* relative to the SSEA1-/cKit- somatic cell controls (Figure 2-S1E). Conversely, somatic gene expression as documented by *Hoxa1* and *Hoxb1* were not enriched in the putative iPGCs relative to somatic cells (Figure 2-S1F).

To test identity of SSEA1+/cKit^{bright} cells derived from V6.5 ESCs we sorted putative iPGCs and cultured them on mouse embryonic fibroblasts (MEFs) supplemented with basic Fibroblast Growth Factor 2 (FGF2), Stem Cell Factor (SCF), Leukemia Inhibitory Factor (LIF) and retinoic acid (RA), a driver of PGC proliferation (Figure 2-1H). This assay has been used previously to confirm PGC identity relative to undifferentiated ESCs, which respond to RA by undergoing differentiation and becoming alkaline phosphatase (AP) negative [15,19,20]. AP+ colony forming ability is almost exclusively associated with SSEA1+/cKit^{bright} population when compared to sorted undifferentiated ESCs, SSEA1+/cKit^{mid} or SSEA1+/cKit^{dim} cells plated at equivalent numbers (Figure 2-1I). The three non-iPGC populations generate mostly mixed colonies or AP negative colonies. Withdrawal of RA, FGF2 and SCF while retaining LIF in the media of RA-treated SSEA1+/cKit^{bright} sorted cells results in the formation of self-renewing pluripotent embryonic germ cells (EGCs), which could be maintained for at least 10 passages (Figure 2-1J).

Day 6 iPGCs have a pre-gonadal, pre-reprogrammed PGC identity

Day 6 EB-derived SSEA1+/cKit^{bright} or Oct4-Gfp+/cKit^{bright} iPGCs consistently express *Dazl* and *Mvh* RNA in addition to *Blimp1* (Figure 2-1F & Figure 2-S1E). Therefore we could hypothesize that putative iPGCs correspond to newly colonized PGCs that have expressed Mvh protein and have potentially undergone whole genome reprogramming. To address this, we performed immunofluorescence for Mvh, which is first detectable in gonadal PGCs at e11.5 [27-29]. We also evaluated DNA demethylation at imprinted and non-imprinted genes, which is erased by e12.5 [30,31]. Immunofluorescence analysis of e10.5 embryos with antibodies against Mvh and Oct4 confirms that e10.5 Oct4-positive PGCs are negative for Mvh protein, whereas gonadal-stage PGCs are Mvh positive (Figure 2-2A). Analysis of SSEA1+/cKit^{bright} sorted iPGCs derived from

V6.5 ESCs reveals that Mvh protein is not detectable above background (Figure 2-2B). We also tested J1 ESC-derived iPGCs and were unable to detect Mvh protein similar to V6.5 iPGCs (data not shown). Furthermore we evaluated H3K27m3 in SSEA1+/cKit^{bright} iPGCs, a histone modification that is depleted from the PGC genome from e11.5-e12.5 [4]. We found a high nuclear content of H3K27me3 in iPGCs (Figure 2-2C). Together this data suggests that the SSEA1+/cKit^{bright} iPGCs are pre-reprogrammed and younger than e11.5.

To further confirm a pre-reprogrammed identity, we next evaluated the methylation status of an imprinted gene (*Snrpn*) and two non-imprinted loci, *Xist* and *Intracisternal A Particle 1 (IAP),* by bisulfite sequencing (Figure 2-2D). Analysis of undifferentiated ESCs shows that the differentially methylated region (DMR) of *Snrpn* is 49.6% methylated, the *Xist* promoter is 66.3% methylated, and *IAP* is 83.5% methylated. In the putative iPGCs, methylation at the *Snrpn* DMR is modestly reduced to 38.7%, while *Xist* and *IAP* methylation levels are the same as ESCs. To determine if the DMR of *Snrpn* also exhibits partial demethylation in endogenous PGCs, we performed bisulfite sequencing of sorted PGCs from e9.5 and e10.5 *Oct4-gfp* embryos. Methylation at the *Snrpn* DMR in e9.5 and e10.5 PGCs from the embryo were still present (47.6% and 54.2% respectively) consistent with previously published findings [31]. Furthermore, we observed evidence of demethylation at the 5' and 3' ends of three clones in endogenous PGCs at e9.5 and two sequences at e10.5 similar to what was observed in iPGCs (Figure 2-2D, arrows). Taken together, using real time PCR, immunofluorescence and bisulfite sequencing, our data strongly argue that the Blimp1-positive PGCs isolated from EBs at day 6 of differentiation correspond to a pre-e11.5 stage germ cell *in vitro*.

Germ line signature genes including *Blimp1* are co-expressed in single SSEA1+/cKit^{bright} cells

To determine whether the relative levels of PGC signature genes in SSEA1+/cKit^{bright} cells are comparable to the levels found in PGCs sorted from the embryo prior to e11.5, we sorted Gfp+ cells from *Oct4-gfp* embryos at 9.5 and e10.5 (Figure 2-3A-C, shown is e9.5). A distinct Gfp+ population was detected from e9.5 to at least e13.5 (Figure 2-3C and data not shown). Sorted Oct4-gfp+ PGCs from the embryo are SSEA1+ and exhibit bright cKit+ staining (Figure 2-3C). We confirmed that the Gfp+ cells are PGCs due to enriched expression of *cKit, Blimp1, Stella,* and *Mvh* relative to the Gfp- somatic cells at a population level by real time PCR (Figure 2-3D). Detection of *cKit* and *Blimp1* RNA in the Gfp- population was not unexpected as these genes are also expressed in endothelial and hematopoietic cells during early embryogenesis [32,33].

We next evaluated the transcriptional identity of undifferentiated ESCs, iPGCs and embryonic PGCs at e9.5 and e10.5 at a single cell level by examining expression of five signature PGC markers (*Blimp1, Stella, Prdm14, Dnd1* and *Dazl)* using the BioMark Fluidigm Real Time PCR platform (Figure 2-3E-H). We evaluated 38 single undifferentiated ESCs (Figure 2-3E), 34 single embryonic Oct4-Gfp+ PGCs at e9.5 (Figure 2-3F), 24 single Oct4-Gfp+ PGCs at e10.5 (Figure 2-3G), and 30 iPGCs from day 6 EBs (Figure 2-3H). In undifferentiated ESCs, 17 of the 38 cells (44%) expressed *Blimp1.* Of the 17 *Blimp1*+ cells, 6 did not express *Stella* and 12 did not express *Dnd1*. In contrast to ESCs where less than 50% of cells expressed *Blimp1*, 100% of e9.5 and e10.5 PGCs from the embryo and iPGCs expressed *Blimp1* (Figure 2-3F-H). Heat maps of the single cell analysis indicate that e9.5 PGCs are relatively homogeneous when comparing individual cells to each other for each gene, whereas at e10.5 and in iPGCs

expression levels between individual cells is more heterogeneous (Figure 2-3G,H). Critically, only one cell in the iPGC cohort was not a germ cell (Figure 2-3H, asterisk).

We next examined expression levels of each gene for all cells that co-expressed *Blimp1*, *Stella, Dnd1,* and *Prdm14* relative to levels in SSEA1+/cKit^{bright} iPGCs. We first compared ESCs to iPGCs and found that of the 4 *Blimp1*+ ESCs that co-expressed *Stella*, *Prdm14* and *Dnd1* (10.5%), the transcript levels were significantly lower than those in iPGCs (Figure 2-3I-L). However, comparison of iPGCs to embryonic e10.5 PGCs revealed no significant difference with regard to *Blimp1, Prdm14,* and *Dnd1* expression levels (Figure 2-3I,K-L). In single cells that also co-expressed *Dazl,* ESCs displayed significantly diminished *Dazl* levels, but no significant difference was found between iPGCs and e10.5 endogenous PGCs (Figure 2-3M). *Stella* levels were statistically different between all groups, with SSEA1+/cKit^{bright} iPGCs on average expressing intermediate levels between e9.5 and e10.5 embryonic PGCs (Figure 2-3J). We propose that the intermediate levels of *Stella* in iPGCs between e9.5 and e10.5 PGCs may indicate that iPGCs are developmentally equivalent to a period of germ cell differentiation between e9.5 and e10.5.

Gene expression profiling by microarray reveals iPGCs repress a mesoderm transcriptional program and identifies a novel marker of in vitro PGC formation from ESCs

Although analysis of five critical PGC-expressed genes at a single cell level was informative for ensuring that >96% of SSEA1+/cKit^{bright} iPGCs have a *Blimp1*+ PGC identity, our next goal was to obtain a more comprehensive transcriptional portrait of iPGCs derived from day 6 EBs by performing microarray analysis using Affymetrix Mouse Genome chips followed by D-Chip

analysis and examine expression of ESC-expressed genes and somatic genes (Figure 2-4A,B). We profiled the SSEA1+/cKit^{bright} fraction from V6.5 EBs at day 6 of differentiation (Samples A and B), Oct4-Gfp+/cKit^{bright} cells from day 6 *Oct4-gfp* EBs (Samples C and D), undifferentiated V6.5 SSEA1+/cKit+ ESCs (Samples E and F), and SSEA1-/cKit- and Oct4-/cKit- day 6 EB cells (Samples G-J).

Cluster analysis of genes that are differentially expressed at greater than three-fold between undifferentiated ESCs and iPGCs (348 genes, $p<0.01$) generated four major transcriptional clusters (Figure 2-4A & Table 2-1). Gene ontology (GO) analysis of Cluster *I* (enriched in ESCs but not iPGCs or somatic cells) identified genes associated with transcription factor activity and DNA binding. Cluster *II (*enriched in iPGCs but not ESCs or somatic cells) revealed enrichment in genes associated with hydrolyase activity, cytoplasmic proteins and MAPK signaling pathways. Genes in Cluster *III* (enriched in ESCs and somatic cells but not iPGCs) were associated with GO terms for stress fibers and actin filament bundle genes. Finally, GO analysis of Cluster *IV* (repressed in iPGCs and ESCs but not somatic cells) revealed genes associated with mesoderm formation including heart and blood development, and morphogenesis. Together, these data suggest that iPGCs repress genes associated with mesoderm differentiation, similar to what has been proposed for endogenous PGC formation through the activity of *Blimp1* [2,7]. Expression of candidate PGC genes from the microarray was validated by real time RT-PCR from Oct4-Gfp+/cKit^{bright} iPGCs (Figure 2-4B). Similarly, expression of additional somatic genes (*Hoxa2* and *Hoxc5)* which were not evaluated earlier (Figure 2-S1F) revealed undetectable expression in iPGCs, whereas somatic cells were positive.

Next, we compared our microarray data between iPGCs and undifferentiated ESCs to identify a

marker that could distinguish between these two cell types. We identified *Inhibitor of DNA binding 4 (Id4)* as being significantly higher in iPGCs relative to ESCs. We confirmed the microarray data showing significant enrichment of *Id4* RNA in independently collected iPGC samples relative to undifferentiated ESCs (Figure 2-4C). To determine if Id4 protein is expressed in iPGCs, we performed immunohistochemistry of day 6 EBs with SSEA1 and Id4, and identified Id4 positive cells within the clusters of SSEA1+ cells (Figure 2-4D). Likewise, immunohistochemistry of e10.5 embryos shows that Id4 protein is expressed in SSEA1+ PGCs (Figure 2-4E, arrow). However, Id4 was also expressed in the surrounding embryonic somatic cells. Taken together, Id4 is a new marker for distinguishing iPGCs from undifferentiated ESCs, but does not distinguish PGCs from somatic cells of the embryo.

Blimp1 is specifically required for iPGC differentiation from EBs

Dosage of *Blimp1* is essential for the specification of PGCs *in vivo* [2,32]. To determine if the emergence of SSEA1+/cKit^{bright} PGCs *in vitro* is similarly dependent upon *Blimp1* expression, we derived *Blimp1*$^{fl/fl}$ ESCs from e3.5 blastocysts. We performed Y chromosome FISH to identify a male line (Figure 2-5A), and generated three independent *Blimp1* knockout sub-lines (*Blimp1*$^{\Delta/\Delta}$) via transfection of Cre recombinase fused to Gfp followed by re-plating of Gfp+ cells at limiting dilutions. Clones were screened by Southern blot to verify *Blimp1* deletion (Figure 2-5B). To compare overall self-renewal and pluripotency in *Blimp1*$^{\Delta/\Delta}$ cells relative to the parental line, we performed flow cytometry for SSEA1 under self-renewing conditions in the presence of LIF (Figure 2-5C), and teratoma analysis by injection of undifferentiated ESCs into the testicles of SCID mice (Figure 2-5D). In both assays, all *Blimp1*$^{\Delta/\Delta}$ lines were indistinguishable from parental *Blimp1*$^{fl/fl}$ cells, indicating that loss of *Blimp1* does not cause gross defects in overall ESC self-renewal or differentiation.

Finally, to evaluate *in vitro* PGC formation, we performed paired differentiation experiments with *Blimp1*$^{fl/fl}$ and *Blimp1*$^{\Delta/\Delta}$ lines and evaluated iPGC differentiation by flow cytometry (Figure 2-5E). Quantification of SSEA1+/cKit^{bright} cells revealed that iPGCs constitute approximately 3-4% of the live cell EB population in the parental *Blimp1*$^{fl/fl}$ line at day 6 (Figure 2-5F). In contrast, all *Blimp1*$^{\Delta/\Delta}$ sub-lines displayed between a 70-90% decrease in SSEA1+/cKit^{bright} iPGCs, with the average percentage constituting less than 1% of the EB in all three sub-lines examined (Figure 2-5F,G). Functionally, this demonstrates that sorting for SSEA1+/cKit^{bright} iPGCs captures a *Blimp1*-dependent PGC population *in vitro*, whereas generation of SSEA1+/cKit^{mid} and SSEA1+/cKit^{dim} fractions of EBs do not exhibit the same reliance on *Blimp1* as *in vivo* PGCs.

DISCUSSION

Emerging cell populations in the early embryo are challenging to investigate. Therefore, we used mouse ESCs from multiple genetic backgrounds to differentiate transgene-free, pre-gonadal stage PGCs where 100% of the single iPGCs express *Blimp1 in vitro.* Here we show that sorting for the cKit^{bright} fraction of SSEA1+ cells at day 6 of differentiation when the population is first discernable yields an iPGC population with an identity suggestive of PGCs younger than e11.5.

One of the major challenges in the ESC and PGC fields has been to distinguish early progenitor PGCs from undifferentiated ESCs due to their similar expression patterns. Indeed, e11.5 PGCs isolated from the genital ridge prior to sex determination cluster very closely to undifferentiated ESCs in 2-dimensional principle component analysis after microarray [11]. Therefore, it has been

proposed that ESCs originate from a progenitor germ cell consistent with detectable expression of PGC-signature genes, including *Dazl* and *Tissue non specific alkaline phosphatase* in the undifferentiated state [34], [35]. Although our studies do not address the origin of ESCs, our data does indicate that a small nascent PGC-like population corresponding to about 10% of cells can be identified in an ESC culture in the self-renewing state, agreeing strongly with previous work which demonstrated that *Dazl* null ESCs exhibit reduced expression of PGC-signature genes [8]. However, our data also show that despite co-expression of germ cell genes in these 10% of cells, the transcript levels are significantly lower than the levels found in *bona fide* PGCs isolated form the embryo between e9.5-e10.5 as well as the iPGCs. Taken together, our data argues that the majority of undifferentiated ESCs are not PGCs, and that a single cell analysis is critical to uncouple differences between ESCs and progenitor PGCs.

In the current study, we identified Id4 as a new marker enriched in iPGCs relative to undifferentiated ESCs. Id4 was recently found to be a germ line marker expressed in gonocytes and spermatogonia of postnatal and adult murine gonads [36]. We extend these findings to show that Id4 is expressed during the earliest stage of germ line development, prior to gonadal colonization (Figure 2-4E). Interestingly, *Id4* similar to *Stella* constitutes a marker for defining PGC identity yet has no functional role in specifying PGC fate [36-38]. However, by combining Id4, SSEA1, and Oct4 expression in day 6 EBs, we propose a model for germ line formation *in vitro* that involves the generation of multiple SSEA1+/Oct4+ niches during EB formation, with Id4+ iPGCs emerging from within these niches (Figure 2-6). We propose that similar to PGC development in the allantois of the embryo, the tight clustering of SSEA1+/Oct4+ cells creates a microenvironment in the EB to protect the iPGCs against somatic cell differentiation signals [39]. Given that Id4+ cells constitute only a subpopulation of cells within SSEA1+ clusters, we hypothesize that the clusters are composed of a heterogeneous mixture of immature cells,

including epiblast-like cells (*Stella* negative, *Blimp1* negative), PGC precursors (*Stella* negative, *Blimp1* positive) and definitive Id4-positive cKit^{bright} iPGCs (Figure 2-6). Whether the SSEA1+/cKit^{bright} PGCs emerge from a subpopulation of the SSEA1+/cKit^{mid} fraction of cells remains to be determined. However, our data strongly argue that iPGCs do not differentiate from SSEA1+/cKit^{dim} cells, which have no colony forming potential, and express high levels of *Hoxa1* and *Hoxb1,* indicating commitment to a somatic fate.

Although our data suggest that the iPGCs are younger than e11.5 of development due to lack of Mvh protein expression, it is conceivable that iPGCs at day 6 are more similar to e11.5 in some aspects, but have not received the appropriate cues to express Mvh protein. The signals that promote Mvh protein expression in PGCs at e11.5 are not well understood, but one study has indicated that gonadal somatic cells are involved in this process [29]. Lack of Mvh protein expression in our model suggests that the hanging drop EB system by day 6 of differentiation does not provide the necessary signals to promote developmental progression to Mvh protein-positive iPGCs. This result implies that progression of iPGCs *in vitro* may require a gonadal niche to promote differentiation to the Mvh protein-positive stage. Indeed, while this manuscript was under review, Hayashi and colleagues demonstrated that a neonatal seminiferous tubule niche was necessary to promote differentiation of ESC-derived PGCs, which this group called PGC like cells (PGCLCs), into functional post-meiotic male germ cells [40]. In these studies, PGCLCs were isolated using SSEA1 and Integrin Beta 3 and were hypothesized to be equivalent to e9.5 of development. Similar to this group, iPGCs isolated at day 6 also express significantly high levels of *Integrin Beta 3* RNA (Figure 2-S2).

In the current study, we successfully acquired PGCs in the *Blimp1*-positive stage of development. *Blimp1* is not expressed in meiotic or post-meiotic cells and therefore our model is not useful for evaluating meiotic progression; however, we propose that this model can be used to successfully evaluate molecular events in PGC formation prior to gonadal colonization, gonadal reprogramming and sex determination. As an example of the utility of our model, one hundred male e10.5 embryos would be required to obtain approximately 100,000 PGCs via FACS, if we estimate that there are ~1,000 PGCs per embryo at this developmental age [5]. In contrast, generating iPGCs equivalent to e9.5-10.5 of development using ESC differentiation required 50 plates of hanging drop EBs, which takes 1 hour to set up from only two wells of undifferentiated ESCs. This yields on average 150,000-175,000 *Blimp1* positive iPGCs at day 6 of differentiation, resulting in more than 100-fold enrichment in cell numbers over embryonic dissections [41].

In conclusion, we propose that the ESC-to-PGC differentiation model is an essential tool for examining molecular events in PGC development. In this study we developed a model that specifically captures the *Blimp1*-positive stage of male PGC formation prior to the expression of Mvh protein. This period of germ cell development (prior to e11.5) is uniquely regulated in mammals and is not conserved with lower model organisms such as *Drosophila*, *C. elegans*, frog, and chick (for example, the role of *Blimp1*). Therefore, creating models that study the initial formation of mammalian PGCs such as the one described here, as well as extending this model to female ESC lines, will be critical to our understanding of the mechanisms that govern fundamental principles of inheritance via the germ line.

MATERIALS AND METHODS

Ethics Statement

Mouse embryo dissection, breeding colony maintenance, and animal surgery were all performed following Institutional Approval for Appropriate Care and use of Laboratory animals by the UCLA Institutional Animal Care and Use Committee (Chancellor's Animal Research Committee (ARC)), Animal Welfare assurance number A3196-01.

Cell Culture and EB Differentiation

All ESC lines in this study were maintained as described previously with lot-tested FBS (Hyclone Lot #ATJ33070) on inactivated CF-1 mouse embryonic fibroblasts (MEFs) [42]. Cells were passaged every three days at 5,300 cells/cm^2. For EB formation, ESCs were subjected to MEF depletion by plating a single cell suspension on tissue culture dishes twice for five minutes each. Cells were seeded in drops of 20 microliters each containing 300 cells on the lids of Petri dishes with 5mL PBS in the plate bottom and cultured in the absence of LIF for six days, with addition of 3.5mL PBS on day 3 of differentiation. For ESC derivation, e3.5 blastocysts were isolated from homozygous *Blimp1*$^{flox/flox}$ (C57BL6/J) crosses and cultured in ESC media containing PD98059 (Cell Signaling) for four days. ESC lines were then passaged and maintained routinely. To generate *Blimp1* null ESCs, *Blimp1*$^{flox/flox}$ cells were transfected with pCAG-Cre:Gfp[43] and sorted to generate sublines.

Mice

Oct4-gfp embryos were dissected and dissociated with TrypLE (Invitrogen) prior to flow cytometry or FACS. For teratoma analysis, 100,000 ESCs were injected into the testicles of SCID recipient mice and collected 6 weeks after transplant for histology [40].

iPGC Colony Assay

iPGCs were sorted from EBs by FACS and re-plated on inactivated CF-1 MEFs. iPGCs were cultured in ESC media supplemented with 15 $ngml^{-1}$ bFGF2 (R&D), 30 $ngml^{-1}$ SCF (Peprotech), and 2 micromolar retinoic acid (Sigma), for five days as described previously [15]. Cells were cultured for five days with daily media changes followed by assaying for AP activity. EGCs were derived from iPGCs by culture of iPGCs for five days with LIF/SCF/bFGF2/RA, followed by passaging in LIF-only containing media for subsequent passages.

Flow cytometry and FACS

Staining for SSEA1 (DSHB, 1:200) and cKit (BD, 1:200) was performed on ice. Indirect labeling was performed with Cy5-conjugated goat anti-mouse IgG and IgM (1:500) and PE-conjugated goat anti-rat IgG (1:1000) (Jackson ImmunoResearch). 7AAD or DAPI were added prior to all acquisitions to examine only live cells for downstream analyses with FlowJo software (TreeStar).

Immunostaining

Embryoid bodies were fixed and embedded in paraffin according to standard protocols. For iPGC stains, cells were sorted by FACS and plated onto poly-lysine coated cover slips. The following antibodies were used at the indicated dilutions: SSEA1 (DSHB, 1:100), Oct4 (1:100, Santa Cruz), Mvh (1:100, Abcam), and H3K27m3 (1:500, Millipore). All samples were incubated with primary antibodies overnight at 4°C. Sections were washed, incubated with FITC anti-mouse IgM, TRITC anti-goat IgG, or FITC/TRITC conjugated anti-rabbit IgG antibodies (Jackson Immunoresearch) for 30 minutes at room temperature. Y chromosome FISH was performed on chromosome spreads. SSEA1 and Id4 immunohistochemical detection was performed using anti-Id4 (1:100, Novus Biologicals) and anti-SSEA1 (DSHB) with standard protocols (Vector Labs).

Real-time PCR

RNA was extracted from sorted samples using the RNEasy Micro Kit (Qiagen) and reverse-transcribed using Superscript RT II (Invitrogen). All gene expression analysis was performed using commercially available TaqMan Gene Expression Assays (Applied Biosystems), with the exception of *Id4,* which as examined by SYBR Green PCR (Roche). See Table 2-2 for additional primer information. CT values were normalized to *Gapdh* expression and expressed as fold change relative to a control cell type referenced in each experiment.

Single-cell Real Time RT-PCR

Single cells were sorted by FACS and subjected to reverse transcription and specific target amplification of relevant genes using the Fluidigm BioMark 48.48 dynamic gene expression system according to manufacturer's instructions, with PCR performed by the UCLA Genotyping and Sequencing core facility. A dilution series of cells were used as detection controls and also to establish primer correlation coefficients and ensure linear amplification of amplicons. Heat map data was generated using Fluidigm Real Time PCR Analysis software.

Microarray and Data Analysis

RNA extraction, labeling amplification and hybridization to Affymetrix Mouse Genome 430 2.0 arrays were performed as previously described [42]. Analysis was performed using model-based expression and invariant set probe normalization using D-Chip software [44]. Gene ontology (GO) terms were identified using DAVID [45,46]. Microarray data is deposited under GEO accession number GSE33121.

Bisulfite Sequencing

Genomic DNA was isolated from sorted samples (Zymo Research). Bisulfite conversion was performed using the EZ DNA Methylation Kit according to manufacturer's instructions (Zymo Research). PCR was performed on bisulfite converted genomic DNA and cloned into pCR2.1-

TOPO (Invitrogen). Clones were sequenced and aligned using Lasergene software (DNASTAR). See Table S2 for PCR primer information.

Southern Blot

Prdm1/Blimp1 deletion was verified with dUTP-digoxigenin-labeled probe generated by PCR upstream of the deleted exons of *Blimp1* (fwd:5'- CTCGTGGCTCTTGTGTGTGT -3', rev:5'- AACGCTGTACCCATGACTCC -3'), after digestion with *EcoRI*. Detection of wild type (15 kb), flox, (13.5 kb), and KO (10 kb) alleles of *Blimp1* have been described [47].

ACKNOWLEDGEMENTS

We acknowledge Konrad Hochedlinger for *Oct4-gfp* mice, Robert Blelloch for V6.5 ESCs, Hong Wu and Rudolph Jaenisch for *Oct4-gfp* ESCs, and Kathrin Plath for *Rosa26-gfp* ESCs. We thank Katrina Adams, Nicole Walsh, Xinmin Li, Bennet Novitch, and Michael Teitell for technical assistance, and Felicia Codrea and Jessica Scholes for FACS. The SSEA1 antibody developed by D. Solter and B. Knowles was obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by the University of Iowa.

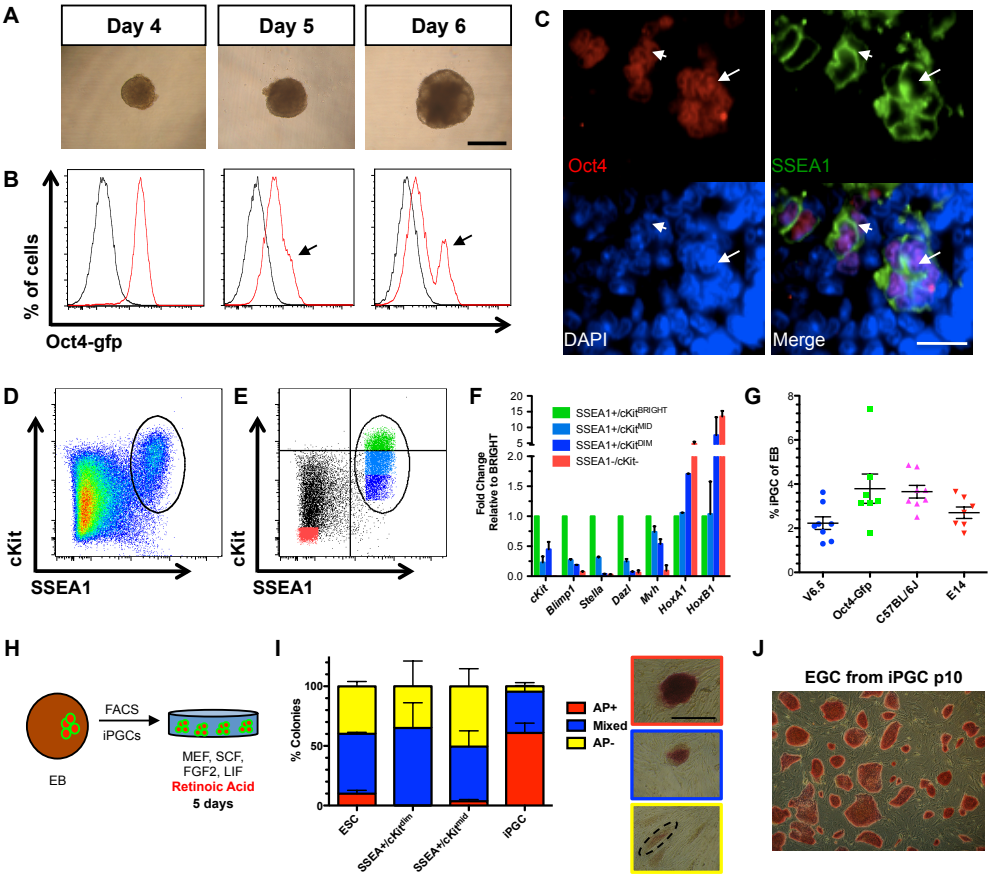

Figure 2-1. Transgene-free method for isolating iPGCs from embryoid bodies.

A: V6.5 embryoid bodies in hanging drops at days 4, 5, and 6 of differentiation. Scale bar=500 microns. B: Oct4-Gfp expression (red) relative to V6.5 EBs (black) at days 4, 5 and 6 of differentiation. Arrows indicate shoulder of Oct4-Gfp^{bright} cells at day 5 and an Oct4-Gfp^{bright} peak at day 6. C: Immunofluorescence of EBs at day 6 for Oct4 (red) and SSEA1 (green). Double positive cells localize in discreet clusters (arrow). Scale bar=20 microns. D: Flow cytometry plot of V6.5 day 6 EBs stained for SSEA1 and cKit. Oval gate defines the SSEA1+/cKit+ side population. E: Flow plot day 6 EBs from V6.5 ESCs fractionated by expression of SSEA1 and cKit into SSEA1+/cKit^{bright} (green), SSEA1+/cKit^{mid} (light blue), SSEA1+/cKit^{dim} (dark blue), and SSEA1-/cKit- cells (red) populations. Quadrant gates are drawn to demonstrate the criteria for selecting SSEA1+/cKit^{bright} cells. The remaining cKit+ population was split into two equal fractions, mid and dim. F: Semi-quantitative real-time PCR from the populations isolated in E, with levels normalized to *Gapdh*. SSEA1+/cKit^{bright} cells are set at 1.0. Data is from two biological replicates each performed in technical duplicate. Error bars represent s.e.m. G: Percentage of live iPGCs acquired from differentiation of ESCs of different genetic backgrounds. Each line was tested at least seven independent times. H: Diagrammatic representation of iPGC replating assay. I: Quantification of alkaline phosphatase (AP) staining of colonies derived from indicated cell populations after 5 days of culture. Right, representative images of colony types. Scale bar=500 microns. J: Self-renewing EGCs at passage 10 derived from RA/FGF2/LIF/SCF cultured iPGCs, followed by routine passaging in the presence of LIF only. Error bars represent s.d.

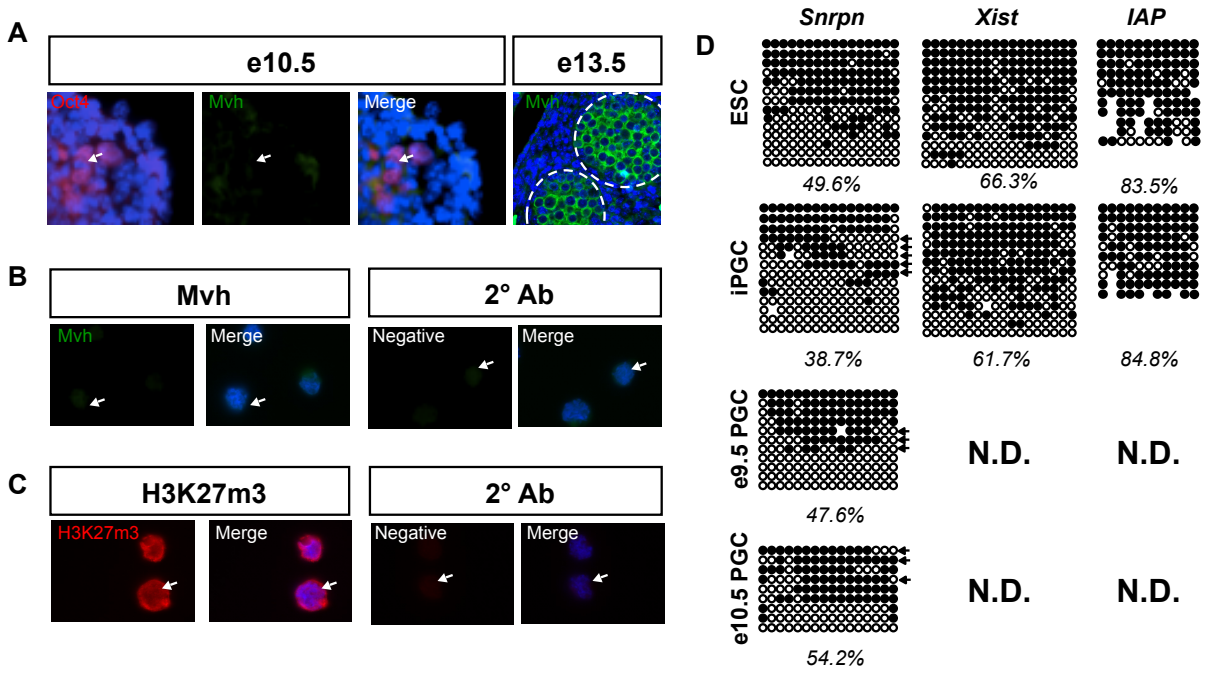

Figure 2-2. iPGCs have characteristics of pre-gonadal, pre-reprogrammed *in vivo* PGCs.

A: Immunofluorescence of pre-gonadal e10.5 PGCs stained for Oct4 (red) and Mvh (green). e13.5 male gonadal PGCs were stained as a positive control. Dotted circles mark the testis cords. B: Sorted SSEA1+/cKit^{bright} iPGCs stained for Mvh (green, left) or secondary antibody alone (right). C: iPGCs stained for H3K27m3 (red, left), and secondary antibody alone (right). Arrows point to individual iPGCs. D: Bisulfite sequencing of ESCs, iPGCs, and endogenous e9.5 PGCs for *Snrpn,* the *Xist* promoter, and *IAP.* Circles represent individual CG dinucleotides, black = methylated and white = unmethylated cytosines. Arrows indicate individual alleles that display characteristic demethylation. N.D. = not determined.

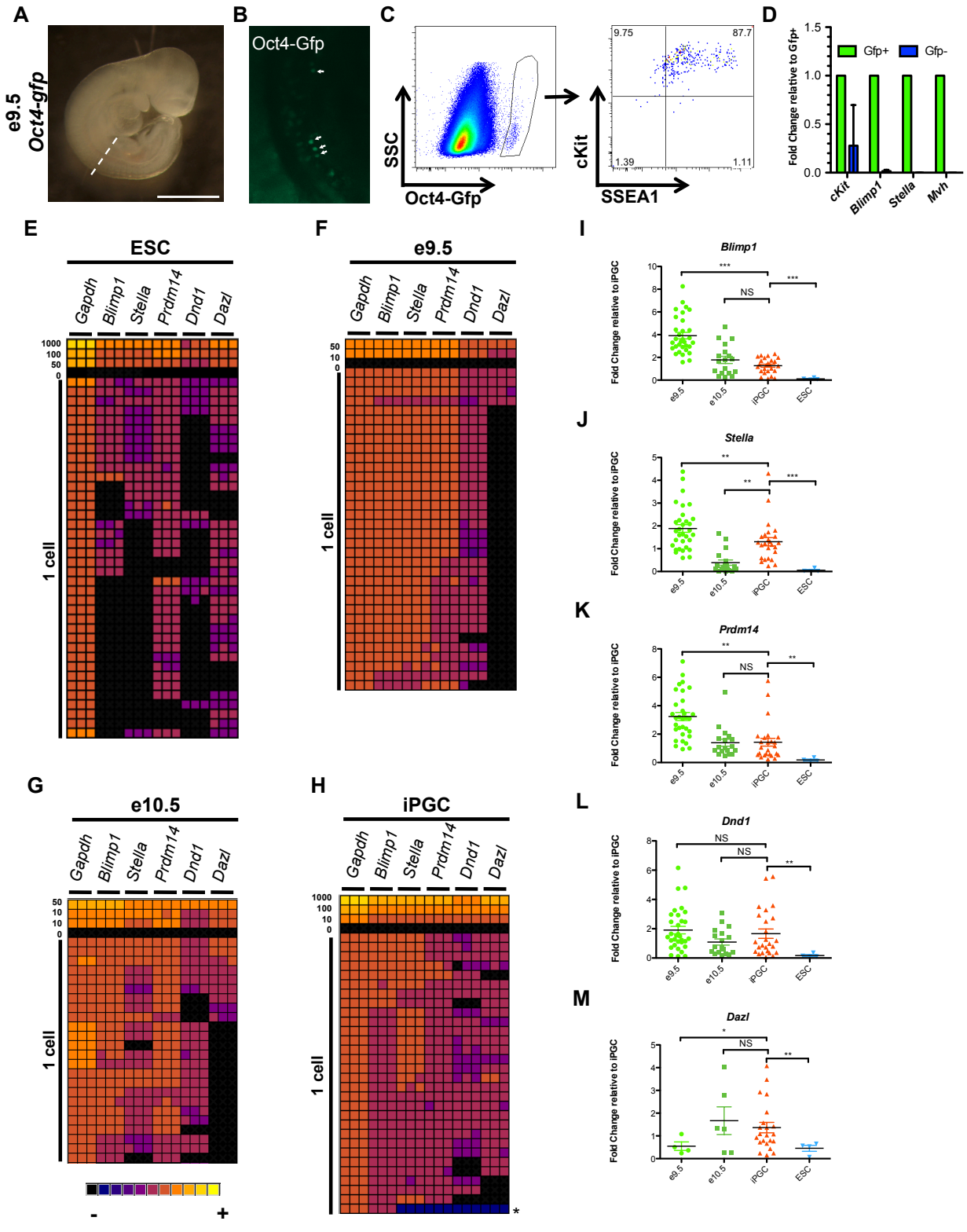
A
e9.5
Oct4-gfp
B
Oct4-Gfp
C
SSC
Oct4-Gfp
9.75
87.7
1.39
1.11
cKit
SSEA1
D
Fold Change relative to Gfp+
Gfp+
Gfp-
cKit
Blimp1
Stella
Myh
E
ESC
F
e9.5
G
e10.5
H
iPGC
Gapdh
Blimp1
Stella
Prdm14
Dnd1
Dazl
1 cell
I
Blimp1
J
Stella
K
Prdm14
L
Dnd1
M
Dazl
Fold Change relative to iPGC
NS
e9.5
e10.5
iPGC
ESC

Figure 2-3. . Developmental staging of pre-gonadal iPGCs at single cell resolution.

A: Bright field image of representative e9.5 *Oct4-gfp* embryo. Dotted line indicates where the embryo was bisected at somite 13 for FACS. Scale bar=1 mm. B: Whole mount confocal microscopy of live embryos with migratory Oct4-gfp+ PGCs within the hindgut (arrows). C: Flow cytometry of the bisected lower half of e9.5 *Oct4-gfp* embryos. Oct4-gfp+ PGCs (circled gate) are also positive for SSEA1 and cKit. D: Real-time RT-PCR of Gfp+ and Gfp- cells. Error bar denotes s.d. E-H: Gene expression analysis at single cell resolution for ESCs (E), e9.5 PGCs (F), e10.5 PGCs (G), and iPGCs (H) represented as a heat map of CT values with expression ranging from not detected (black) to high (yellow). A cell titration was performed as a control to ensure linear amplification of each primer set. Each cell was evaluated for the expression of each gene in technical triplicate. I-M: Semi-quantitative analysis of single cell real time PCR in E-H of cells that co-express *Blimp1, Stella, Prdm14,* and *Dnd1* expressed relative to the average delta CT expression level for each gene in single iPGCs. *p< 1e-03, **p<1 e-04, ***p<1 e-06, NS=not significant.

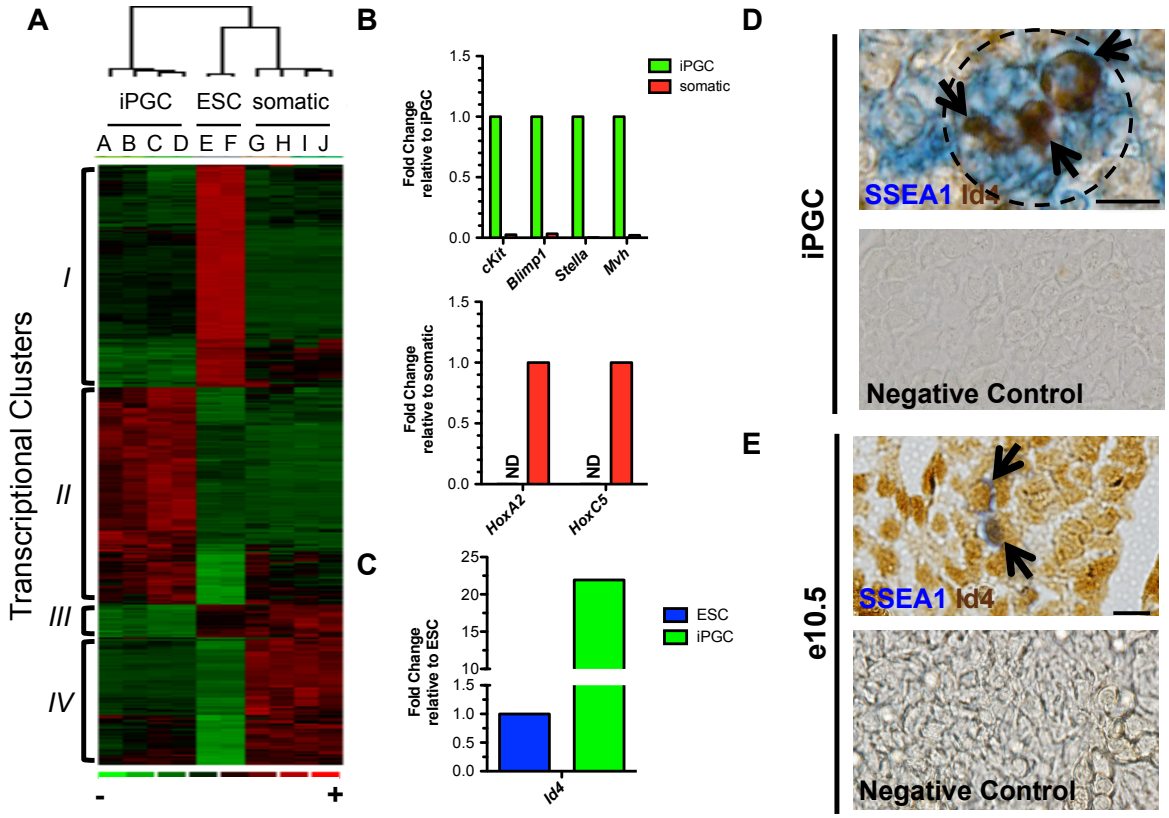

Figure 2-4. Transcriptional profiling demonstrates a PGC program and identifies novel markers for *bona fide* iPGCs from ESCs.

A: Microarray analysis comparing iPGCs (A,B,C,D), V6.5 ESCs (E and F), and the somatic cells of the EB (G through J). Genes differentially expressed by threefold between undifferentiated ESCs and iPGCs are shown ($p<0.01$). Red indicates significant up-regulation, green repression and black no change. **B:** Semi-quantitative RT-PCR validation of microarray in Oct4-Gfp/cKit^{bright} iPGCs versus Gfp-/cKit- somatic cells. ND=no transcript detected. **C:** Real time RT-PCR of *Id4* in V6.5 ESCs (set to 1.0) compared to iPGCs. **D:** Immunohistochemistry of day 6 V6.5 EBs for SSEA1 (blue) and Id4 (brown, arrows). Dotted black line denotes SSEA1+ cluster within the EB. Scale bar=10 microns. **E:** Immunohistochemistry of e10.5 embryos. Id4 is expressed in SSEA1+ PGCs (arrows) and somatic cells. Scale bar=20 microns. Negative controls were performed with secondary antibodies alone.

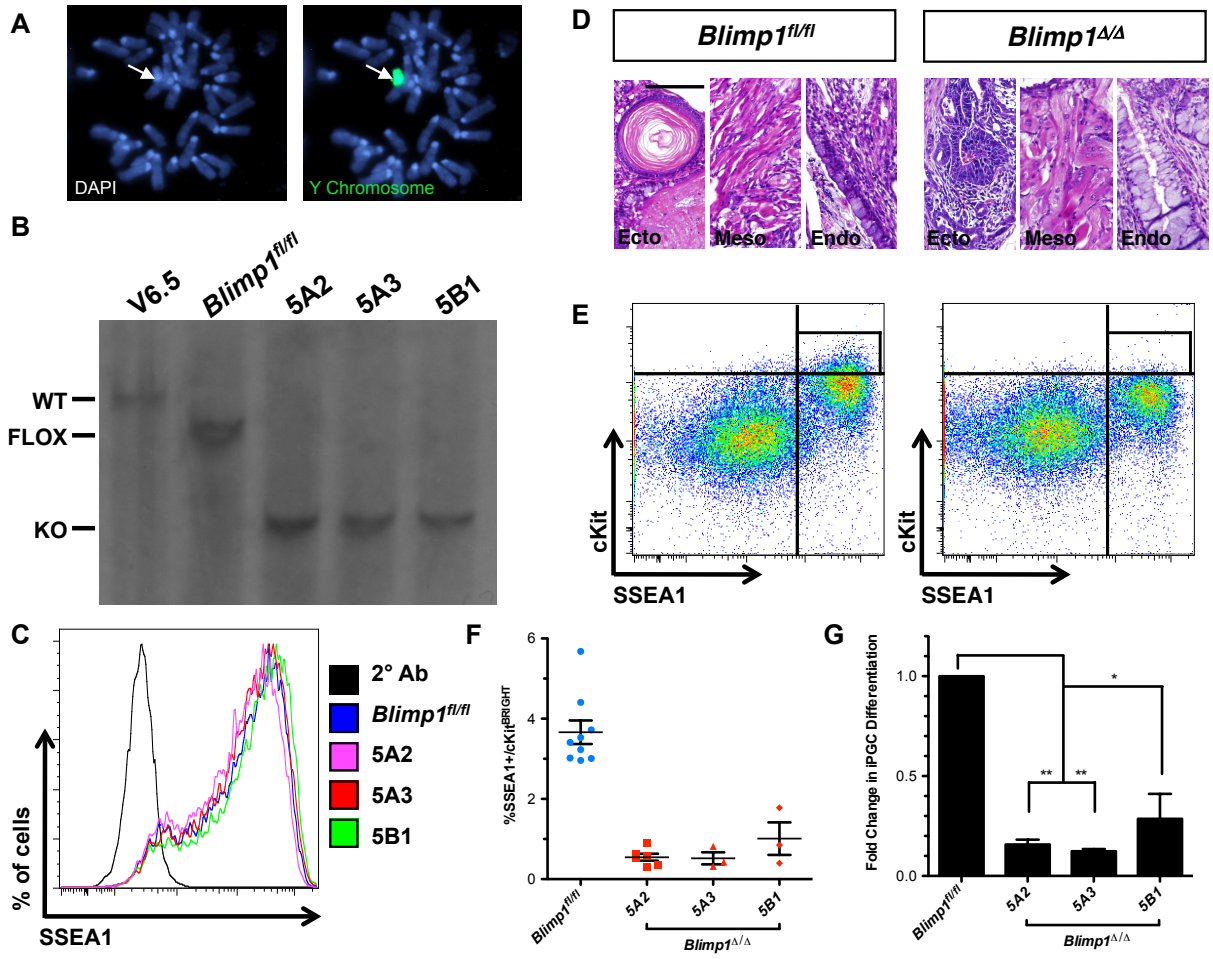

Figure 2-5. *Blimp1* is required for the differentiation of iPGCs from ESCs.

A: DNA-FISH for the Y chromosome in *Blimp1*$^{fl/fl}$ ESCs. B: Southern blot for detection of wild type (WT), flox, and knock-out (KO) alleles of *Blimp1*. C: Flow cytometry for SSEA1 on undifferentiated ESCs. D: Representative histological sections from *Blimp1*$^{fl/fl}$ and *Blimp1*$^{\Delta/\Delta}$ teratomas. All lines were capable of differentiation to ectoderm (Ecto), mesoderm (Meso) and endoderm (Endo). Scale bar=100 microns. E: Representative paired EB differentiations of *Blimp1*$^{fl/fl}$ and *Blimp1*$^{\Delta/\Delta}$ ESCs. Quadrant gates indicate criteria for gating SSEA1+/cKit^{bright} iPGCs, which are contained within the rectangular gate (black lines). F: Percentage iPGC yield in the control *Blimp1*$^{fl/fl}$ line and *Blimp1*$^{\Delta/\Delta}$ sub-lines. Error bars represent s.e.m. G: Quantification of data from F, expressed as a percent of the *Blimp1*$^{fl/fl}$ iPGC yield from each paired experiment. Error bars represent the standard error of the mean. *p< 0.05, ** p< 1x10^{-7}.

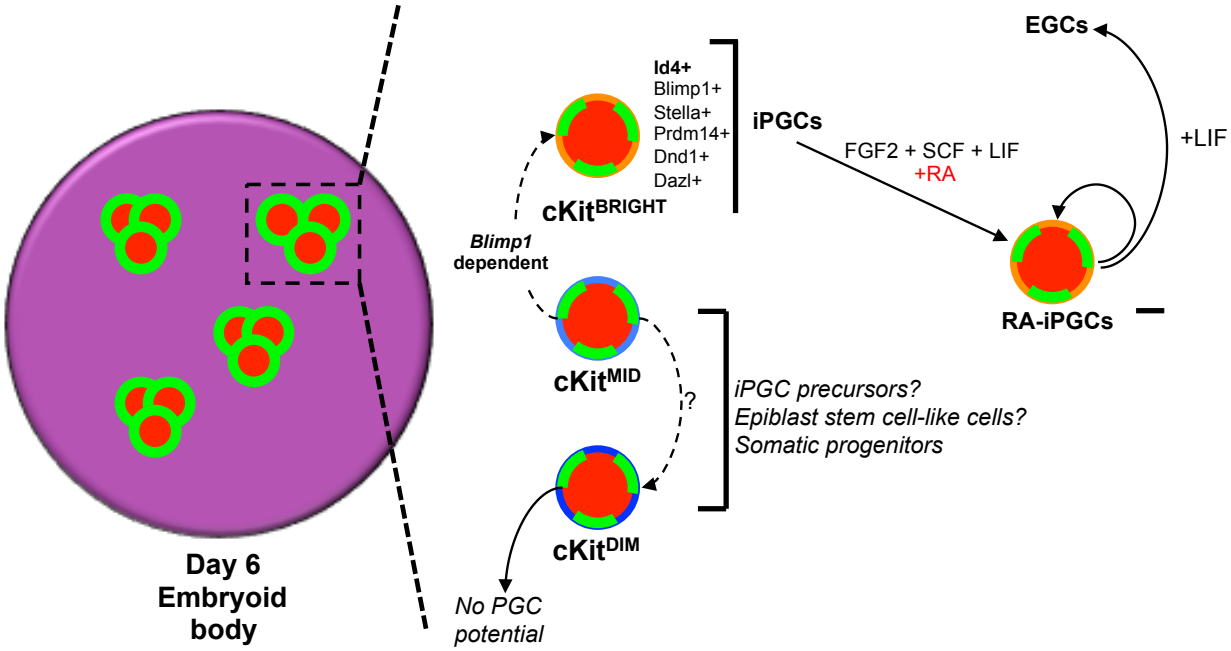

Figure 2-6. Model for iPGC emergence from SSEA1/Oct4+ clusters in EBs.

In day 6 of EBs (pink circle) multiple discreet clusters of Oct4+ (red) and SSEA1+ (green) cells are identified. SSEA1+ cells within these clusters exhibit a range of cKit signal intensities identified by flow cytometry including cKit^{bright} (iPGCs), cKit^{mid} (iPGC precursors and epiblast stem cells) and cKit^{dim} (somatic lineage primed epiblast) cells (black box). Definitive iPGCs are enriched in the cKit^{bright} fraction of SSEA1+ or Oct4+ cells, and the generation of this population *in vitro* is dependent upon *Blimp1*. Using a differential colony forming assay in the presence FGF2, SCF, LIF and RA which promotes survival and proliferation of PGCs, we show that RA-iPGC potential is highest in the cKit^{bright} fraction and is absent in the cKit^{dim} subpopulation of SSEA1+ cells. Furthermore, converting RA-iPGCs to media containing LIF supports the generation of self-renewing EGCs *in vitro*.

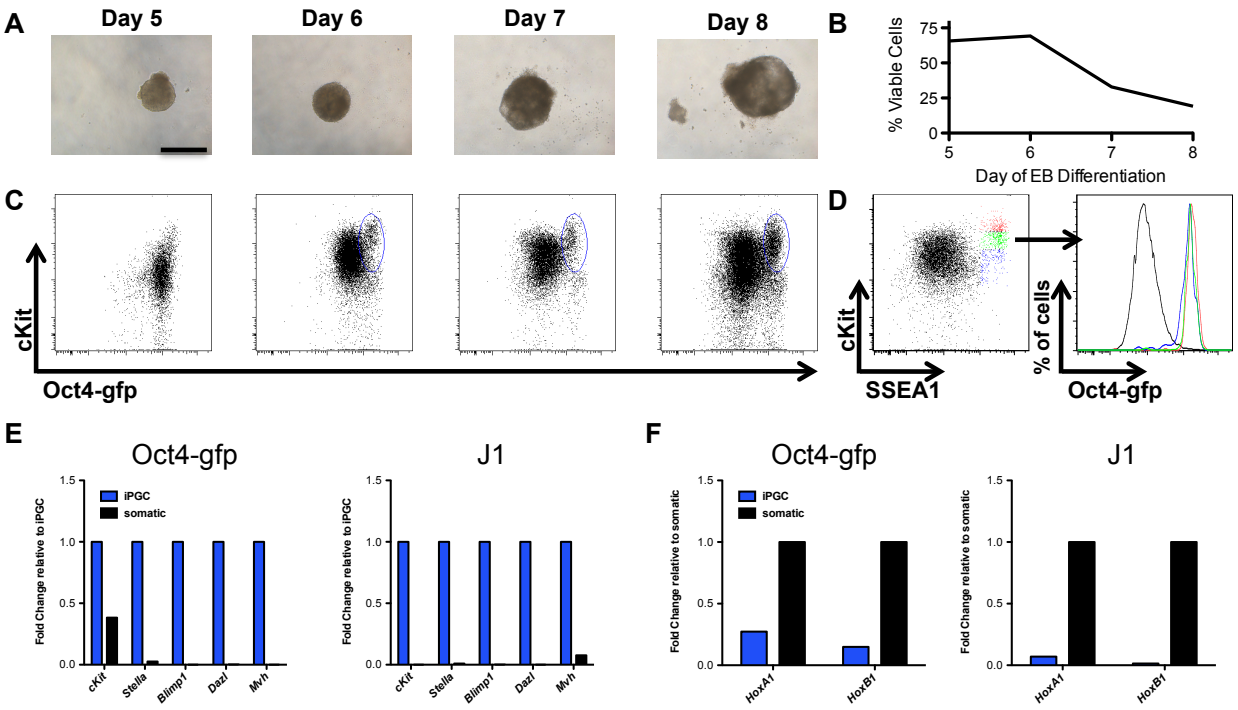

Figure 2-S1. Kinetics of EB formation and the transcriptional identity of iPGCs.

A: *Oct4-gfp* embryoid bodies at days 5-8 of differentiation. Scale bar=500 microns. **B**: Quantification of EB cell viability recorded as the percent of 7AAD- cells at each time point by flow cytometry. **C**: Flow cytometry of the live cell EB fraction for Oct4-gfp and cKit at the corresponding time point. Blue oval indicates the Oct4-gfp+/cKit+ side population, which first appears at day 6. Oct4-gfp+/cKit^{bright} cells correspond to iPGCs. **D**: *Oct4-gfp* EBs at day 6 were stained with SSEA1 and cKit, and Oct4-gfp expression was examined in SSEA1+/cKit^{bright}, SSEA1+/cKit^{mid} and SSEA1+/cKit^{dim} populations. **E**: PGC gene expression data for *Oct4-gfp* and J1-derived iPGCs and somatic cells. **F**: Somatic gene expression data for *Oct4-gfp* and J1-derived iPGCs and somatic cells.

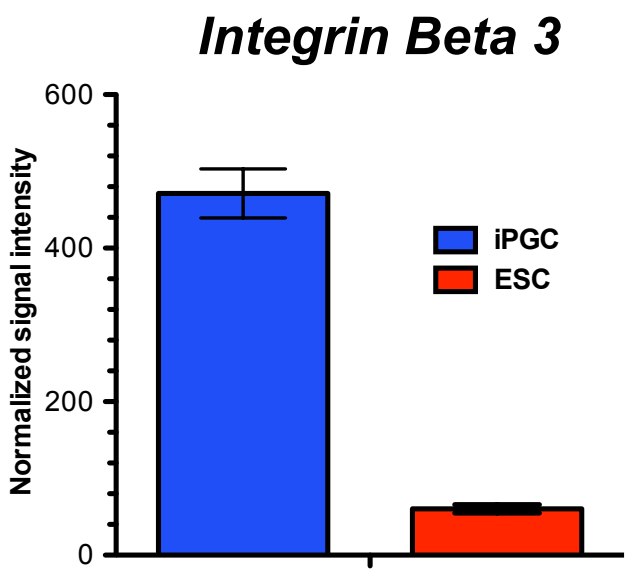

Figure 2-S2. ***Integrin Beta 3*** **is enriched in iPGCs.**

Normalized signal intensity from probe sets for *Integrin beta 3 (Itgb3)* for ESCs and iPGCs were determined from the microarray presented in Figure 2-4A.

Cluster	Affymetrix probe sets	Top GO category	P-Value	Gene Examples
I	126	Transcription factor activity	3.2E-05	*Socs3, Sfrp1, Klf5, Klf4, Etv4, Klf9, Ets2, Hesx1, Hoxc13, Fgf17*
II	126	Cytoplasm	8.0E-03	*Rhox6, Rhox9, Tcl2, Dmrt3, Tex19.2, Map4k2, Id4*
III	19	Stress fiber	2.5E-02	*Fhl3, Anxa2, Parvb,Glipr2, lmna, Itga5*
IV	74	Heart development	4.90E-07	*Tbx20, Cfc1, Fgf15, Msx1, Id2, Id3, Hand1, Wnt3, Wnt5b,Evx1,*

Table 2-1. DAVID Gene Ontology analysis of transcriptional clusters following D-Chip analysis comparing undifferentiated ESCs to iPGCs. ($p<0.001$ and >3-fold difference)

Gene	Primer	Sequence
Gapdh (RT)	FP	ACCACAGTCCATGCCATCAC
	RP	TCCACCACCCTGTTGCTGTA
Id4 (RT)	FP	GAGACTCACCCTGCTTTGCT
	RP	AGAATGCTGTCACCCTGCTT
Snrpn outer (BS)	FP	TATGTAATATGATATAGTTTAGAAATTAG
	RP	AATAAACCCAAATCTAAAATATTTTAATC
Snrpn inner (BS)	FP	AATTTGTGTGATGTTTGTAATTATTTGG
	RP	ATAAAATACACTTTCACTACTAAAATCC
IAP (BS)	FP	TTGTGTTTTAAGTGGTAAATAAATAATTTG
	RP	CAAAAAAAACACACAAACCAAAAT
Xist outer (BS)	FP	TGGTTTGTTTAAGTAGAAGATATATTG
	RP	AAAAATCTTACCAAAACATATCAAAAC
Xist inner (BS)	FP	GTATAGATAGGTGTGTGATTTAATG
	RP	TTTAATATATTTTCTTAAATAAACC

RT=RT-PCR, BS=bisulfite sequencing-PCR, FP=forward primer, RP=reverse primer

Table 2-2. Primers used in this study.

REFERENCES

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

2 Ohinata, Y. *et al.* Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213, doi:10.1038/nature03813 (2005).

3 Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**, 440-458, doi:10.1016/j.ydbio.2004.11.025 (2005).

4 Seki, Y. *et al.* Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134**, 2627-2638, doi:10.1242/dev.005611 (2007).

5 Tam, P. P. & Snow, M. H. Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. *J Embryol Exp Morphol* **64**, 133-147 (1981).

6 Ancelin, K. *et al.* Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat Cell Biol* **8**, 623-630, doi:10.1038/ncb1413 (2006).

7 Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* **22**, 1617-1635, doi:10.1101/gad.1649908 (2008).

8 Haston, K. M., Tung, J. Y. & Reijo Pera, R. A. Dazl functions in maintenance of pluripotency and genetic and epigenetic programs of differentiation in mouse primordial germ cells in vivo and in vitro. *PLoS ONE* **4**, e5654, doi:10.1371/journal.pone.0005654 (2009).

9 Hübner, K. *et al.* Derivation of oocytes from mouse embryonic stem cells. *Science* **300**, 1251-1256, doi:10.1126/science.1083452 (2003).

10 Imamura, M. *et al.* Induction of primordial germ cells from mouse induced pluripotent stem cells derived from adult hepatocytes. *Mol. Reprod. Dev.* **77**, 802-811, doi:10.1002/mrd.21223 (2010).

11 Sabour, D. *et al.* Identification of genes specific to mouse primordial germ cells through dynamic global gene expression. *Hum Mol Genet*, doi:10.1093/hmg/ddq450 (2010).

12 Young, J. C., Dias, V. L. & Loveland, K. L. Defining the window of germline genesis in vitro from murine embryonic stem cells. *Biology of reproduction* **82**, 390-401, doi:10.1095/biolreprod.109.078493 (2010).

13 Payer, B. *et al.* Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. *Genesis* **44**, 75-83, doi:10.1002/gene.20187 (2006).

14 Wei, W. *et al.* Primordial germ cell specification from embryonic stem cells. *PLoS ONE* **3**, e4013, doi:10.1371/journal.pone.0004013 (2008).

15 West, J. A. *et al.* A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* **460**, 909-913, doi:10.1038/nature08210 (2009).

16 Nicholas, C. R. *et al.* Characterization of a Dazl-GFP germ cell-specific reporter. *Genesis* **47**, 74-84, doi:10.1002/dvg.20460 (2009).

17 Mizuno, K. *et al.* Genes associated with the formation of germ cells from embryonic stem cells in cultures containing different glucose concentrations. *Mol. Reprod. Dev.* **73**, 437-445, doi:10.1002/mrd.20395 (2006).

18 Toyooka, Y., Tsunekawa, N., Akasu, R. & Noce, T. Embryonic stem cells can form germ cells in vitro. *Proc Natl Acad Sci USA* **100**, 11457-11462, doi:10.1073/pnas.1932826100 (2003).

19 Geijsen, N. *et al.* Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* **427**, 148-154, doi:10.1038/nature02247 (2004).

20 Qing, T. *et al.* Induction of oocyte-like cells from mouse embryonic stem cells by co-culture with ovarian granulosa cells. *Differentiation* **75**, 902-911, doi:10.1111/j.1432-0436.2007.00181.x (2007).

21 Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-324, doi:10.1038/nature05944 (2007).

22 Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y. & Saitou, M. Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. *Biology of reproduction* **75**, 705-716, doi:10.1095/biolreprod.106.053686 (2006).

23 Yoshimizu, T. *et al.* Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev Growth Differ* **41**, 675-684 (1999).

24 Fox, N., Damjanov, I., Martinez-Hernandez, A., Knowles, B. B. & Solter, D. Immunohistochemical localization of the early embryonic antigen (SSEA-1) in postimplantation mouse embryos and fetal and adult tissues. *Developmental Biology* **83**, 391-398 (1981).

25 Hayashi, K. & Surani, M. A. Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. *Development* **136**, 3549-3556, doi:10.1242/dev.037747 (2009).

26 Morita-Fujimura, Y., Tokitake, Y. & Matsui, Y. Heterogeneity of mouse primordial germ cells reflecting the distinct status of their differentiation, proliferation and apoptosis can be classified by the expression of cell surface proteins integrin alpha6 and c-Kit. *Dev Growth Differ* **51**, 567-583, doi:10.1111/j.1440-169X.2009.01119.x (2009).

27 Fujiwara, Y. *et al.* Isolation of a DEAD-family protein gene that encodes a murine homolog of Drosophila vasa and its specific expression in germ cell lineage. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12258-12262 (1994).

28 Tanaka, S. S. *et al.* The mouse homolog of Drosophila Vasa is required for the development of male germ cells. *Genes & Development* **14**, 841-853 (2000).

29 Toyooka, Y. *et al.* Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mechanisms of Development* **93**, 139-149 (2000).

30 Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-881, doi:10.1038/nature06714 (2008).

31 Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23 (2002).

32 Vincent, S. D. *et al.* The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* **132**, 1315-1325, doi:10.1242/dev.01711 (2005).

33 Yokomizo, T., Ng, C. E., Osato, M. & Dzierzak, E. Three-dimensional imaging of whole midgestation murine embryos shows an intravascular localization for all hematopoietic clusters. *Blood* **117**, 6132-6134, doi:10.1182/blood-2011-02-334037 (2011).

34 Zwaka, T. P. & Thomson, J. A. A germ cell origin of embryonic stem cells? *Development* **132**, 227-233, doi:10.1242/dev.01586 (2005).

35 Chu, L. F., Surani, M. A., Jaenisch, R. & Zwaka, T. P. Blimp1 expression predicts embryonic stem cell development in vitro. *Curr Biol* **21**, 1759-1765, doi:10.1016/j.cub.2011.09.010 (2011).

36 Oatley, M. J., Kaucher, A. V., Racicot, K. E. & Oatley, J. M. Inhibitor of DNA binding 4 is expressed selectively by single spermatogonia in the male germline and regulates the self-renewal of spermatogonial stem cells in mice. *Biol Reprod* **85**, 347-356, doi:10.1095/biolreprod.111.091330 (2011).

37 Ohinata, Y. *et al.* Germline recruitment in mice: a genetic program for epigenetic reprogramming. *Ernst Schering Research Foundation workshop*, 143-174 (2006).

38 Saitou, M., Barton, S. C. & Surani, M. A. A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300, doi:10.1038/nature00927 (2002).

39 Maye, P., Becker, S., Kasameyer, E., Byrd, N. & Grabel, L. Indian hedgehog signaling in extraembryonic endoderm and ectoderm differentiation in ES embryoid bodies. *Mech Dev* **94**, 117-132 (2000).

40 Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532, doi:10.1016/j.cell.2011.06.052 (2011).

41 Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101-1105, doi:10.1038/nature08829 (2010).

42 Lindgren, A. G. *et al.* Loss of Pten causes tumor initiation following differentiation of murine pluripotent stem cells due to failed repression of Nanog. *PLoS One* **6**, e16478, doi:10.1371/journal.pone.0016478 (2011).

43 Matsuda, T. & Cepko, C. L. Controlled expression of transgenes introduced by in vivo electroporation. *Proc Natl Acad Sci USA* **104**, 1027-1032, doi:10.1073/pnas.0610155104 (2007).

44 Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 31-36, doi:10.1073/pnas.011404098 (2001).

45 Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).

46 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

47 Shapiro-Shelef, M. *et al.* Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* **19**, 607-620 (2003).

CHAPTER 3

STAGE-SPECIFIC ROLES FOR TET1 AND TET2 IN DNA DEMETHYLATION IN PRIMORDIAL GERM CELLS

Stage-specific roles for Tet1 and Tet2 in DNA demethylation in primordial germ cells

John J. Vincent[1-3], Yun Huang[6], Pao-Yang Chen[2-3], Suhua Feng[2-3], Joseph H. Calvopiña[2-3], Kevin Nee[2], Serena A. Lee[2], Thuc Le[4-5], Alexander J. Yoon[1,5], Kym Faull[1,5], Guoping Fan[1,3,4], Anjana Rao[6], Steven E. Jacobsen[1-3,7], Matteo Pellegrini[1-3], Amander T. Clark[1-3]*

1 Molecular Biology Institute, 2 Molecular Cell and Developmental Biology, 3 Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, 4 Human Genetics, 5 Pasarow Mass Spectrometry Lab, University of California, Los Angeles, 90095, 6 La Jolla Institute for Allergy and Immunology, La Jolla, CA, 7 Howard Hughes Medical Institute

ABSTRACT

Primordial germ cells (PGCs) undergo dramatic rearrangements to their methylome during embryogenesis, including initial genome-wide DNA demethylation that establishes the germ line epigenetic ground state. The role of the 5mC dioxygenases Tet1 and Tet2 in the initial genome wide DNA demethylation process has not been examined directly. Using PGCs differentiated from either control or *Tet2-/-; Tet1* knockdown embryonic stem cells (ESCs), *in vitro* (iPGCs), we show that iPGC formation and genome-wide DNA demethylation are unaffected by the absence of Tet1 and Tet2, and thus 5hmC. However, numerous promoters and gene bodies were hypermethylated in the mutant iPGCs, consistent with a role for 5hmC as an intermediate in locus-specific demethylation. Together, our results support a revised model of PGC DNA demethylation in which the first phase of comprehensive 5mC loss does not involve 5hmC, and Tet1 and Tet2 instead have a locus-specific role in shaping the PGC epigenome during subsequent development.

INTRODUCTION

DNA methylation is an epigenetic mark involving the addition of a methyl group to the fifth carbon of a cytosine base (5mC). In mammalian cells, DNA methylation is established and maintained mostly in CG sequence contexts, and the amount of cytosine methylation in a given genome is relatively stable [1]. Despite this stability, there are periods in embryonic development where DNA methylation is significantly reduced, including after oocyte fertilization, during pre-implantation embryo development, and during primordial germ cell (PGC) formation [2-13]. Recent work has revealed a critical role for oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) by *Tet methylcytosine dioxygenase 1* (*Tet1*) and *Tet2* in locus specific DNA demethylation in PGCs [4, 14]. However, a role for 5hmC in the initial global depletion of DNA demethylation in PGCs has not been addressed (Figure 3-1A).

PGCs are the founder cells of the metazoan germ line, and abnormal PGC development causes infertility or cancer. Mammalian PGCs are specified *de novo* each generation from the epiblast [15-18], and begin as highly methylated cells [12]. Although methylation is critical for lineage specialization [19], it poses an inherent problem for PGCs, which become globally depleted of DNA methylation by e13.5 [3, 8, 10, 20]. This differential has lead to a long-standing hypothesis that loss of methylation in PGCs prior to e13.5 is necessary to restore the germ line epigenetic ground state, and the mechanisms for this are not well understood [21].

PGCs undergo DNA demethylation in two phases [11] (Figure 3-1A). The first phase involves widespread (global) depletion of cytosine methylation with retention of locus-specific methylation at imprinting control centers (ICCs), single copy genes and repetitive elements [3, 4, 6, 12, 22, 23]. The second phase occurs from e9.5 to e13.5 where methylation is depleted from the PGC genome

in a locus-specific manner [3-7, 10, 11, 14]. Recent work has revealed a role for Tet1 in locus-specific demethylation of meiotic genes [14]. However, given that 5hmC levels were reduced by only 45% in this model it is conceivable that a second Tet protein may have compensated for loss of Tet1. More recently, a double knockdown of *Tet1* and *Tet2* in ESC-derived PGCs revealed a role for Tet1 and Tet2 in demethylation of germ-line genes *Deleted in azoospermia like (Dazl)*, *Maelstrom (Mael)* and *Synaptonemal complex protein 3 (Sycp3)* [4]. However, it was not determined whether Tet1 and Tet2 act to regulate global DNA demethylation in phase 1.

In the current study our goal was to evaluate the role of Tet1 and Tet2 in genome-wide DNA demethylation using the differentiation of PGCs *in vitro* (iPGCs) from embryonic stem cells (ESCs). It has previously been reported that this method robustly captures immature PGCs transcriptionally younger than e10.5 of development at high purity [24]. However it is not known whether global DNA demethylation occurs in this model. Therefore, the goals of the study were two fold. The first was to examine whether differentiation of iPGCs from ESCs involves a genome-wide depletion of DNA methylation from the iPGC genome, and if so, to use this model to determine whether Tet1 and Tet2 regulate the phase 1 genome-wide demethylation in PGCs.

RESULTS

PGCs undergo DNA demethylation in two phases [11]. In phase 1, 5mC is depleted globally from the genome with rare locus-specific retention of methylation including the ICC of *Snrpn* (Figure 3-S1A-D). To determine whether ESC derived PGCs *in vitro* (iPGCs) undergo genome-wide demethylation we used two independently derived ESC lines (V6.5 and *Rosa26-GFP*) and differentiated iPGCs using embryoid body (EB) differentiation. The iPGCs were sorted on day 6 using surface markers stage specific embryonic antigen 1 (SSEA1) and cKit and gating on the

SSEA1+/cKit^{bright} population (Figure 3-S1F). The transcriptional identity of iPGCs was confirmed using single-cell gene expression analysis of forty SSEA1+/cKit^{bright} cells from the XY V6.5 background (Figure 3-1B). In this study the higher the "cross threshold" (Ct), the lower the gene expression with black indicating no detectable Ct, therefore no expression. We found that 38/40 iPGCs co-expressed PGC-genes *Blimp1* and *Dppa3,* with heterogeneous levels of the gene *Dead end 1* (*Dnd1*) as previously reported [24]. We also determined that the XY iPGCs do not express the spermatogonial marker *Plzf*, and are negative for somatic lineage markers *Hoxa1* and *Hoxb1.* Using bisulfite (BS) treatment of iPGC DNA followed by PCR amplification of the *Snprn* ICC, we demonstrate that iPGCs on day 6 are methylated indicating that iPGCs have not completed phase 2 demethylation (Figure 3-S1E).

Next, we performed Whole Genome BS Sequencing (BS-Seq) to compare cytosine methylation in ESCs and iPGCs (Table 3-1). Notably, BS treatment does not distinguish between 5mC and 5hmC [25]. Therefore the use of BS is detecting the sum of 5mC and 5hmC. Cytosine methylation was mapped using BS Seeker [26] with mouse genome build mm9 (UCSC Genome browser) allowing up to 3 mismatches. Using this approach we quantified a statistically significant ($p<0.05$) reduction in the levels of CG methylation in iPGCs relative ESCs (Figure 3-1C). Specifically we determined that on average 75% of cytosines in a CG sequence context in ESCs were methylated, whereas in iPGCs this was reduced to 47%. Cytosine methylation was also observed in non-CG contexts. However, the amounts of non-CG methylation were low, at around 2% or less (Table 3-1). Though non-CG methylation in iPGCs trended towards depletion, this trend did not reach statistical significance.

To map genomic regions where loss of CG methylation occurred, we sequenced undifferentiated ESCs to 6.8X and iPGCs to 6.9X coverage per strand resulting in 478,482,437 cytosines covered ≥4 times in both samples. Sites with delta methylation levels ≥30% were subject to two-way binomial tests, which yielded 11,994,107 CG sites for further analysis. We determined that 8,623,115 methylated cytosines in a CG sequence context were significantly decreased in iPGCs where as only 81,884 methylated cytosines in a CG sequence context were significantly increased in iPGCs relative to ESCs (FDR<= 5%). Chromosomal views of in 1M bp windows revealed a chromosome-wide depletion of CG methylation across all chromosomes in iPGCs (Figure 3-S1G). Metaplots of reference genes showed typical depletion of CG methylation at the transcription start site (TSS) in both iPGCs and ESCs, but general hypomethylation in iPGCs across all upstream and downstream regions (Figure 3-1D). Examination of repeat regions including short and long interspersed nuclear elements (SINEs and LINEs), revealed a similar depletion in CG methylation (Figure 3-S1H,I). Next we used a Shannon Entropy calculation to capture the heterogeneity of methylated cytosines between samples (Figure 3-S1J). We found that the entropy was considerably higher in iPGCs compared to ESCs. This was true for genes, pseudogenes, exons and introns. The one exception was gene promoters (defined as -800bp to +200bp of the TSS), where the entropy was almost equivalent between ESCs and iPGCs. Together, we conclude that loss of cytosine methylation from iPGCs occurred genome-wide and the increased entropy indicates that the iPGC population is heterogeneously (not synchronously) undergoing demethylation.

To determine whether global changes in cytosine methylation correlate with global changes in gene expression, we plotted differentially methylated CG sites (DMS) against the average change in gene expression between ESCs and iPGCs for each reference gene (Figure 3-1E). ESC and iPGC gene expression data was obtained from [24]. Performing a Pearson Correlation

Coefficient for all comparisons revealed that the DMS at promoters, gene bodies, exons and introns exhibited no correlation to either an increase or a decrease in gene expression in iPGCs relative to ESCs (Figure 3-1E, Figure 3-S1K). Unlike the majority of the PGC genome, analysis of CG Islands (CGIs) revealed no change in the percentage of CG methylation (Figure 3-1F). However, a small number of promoter CGIs significantly gained methylation in iPGCs, and these were enriched in gene ontology groups associated with meiosis (Figure 3-1G). This was confirmed by bisulfite-PCR of the meiotic gene *Tex12* CGI (Figure 3-1H). Together these data demonstrate that methylation at CGIs undergoes dynamic and unique reorganization with iPGC differentiation.

We next evaluated the changes in the distribution of methylated CGs in ESCs and iPGCs by mapping methylation levels from 0 to 100 as a fraction of total cytosine methylation (Figure 3-1I). Cytosine methylation in ESCs exhibit a typical bimodal distribution where 59.6% of cytosines had $\geq$80% methylation (red, high), while 12.2% of cytosines had $\leq$20% methylation (blue, low). In contrast, in iPGCs we observed a substantial loss of cytosine methylation from the high category and a near doubling of cytosines in the low category (from 12.2% to 21.6%). The largest change between ESCs and iPGCs was progression to the intermediate category (yellow, >0.2 and <0.8), which more than doubled (from 28.2% to 65.3%).

Finally, we evaluated the symmetry of cytosine methylation (Figure 3-1J). In this analysis, mC+mG refer to symmetrically methylated CG sites where the cytosine from both strands of DNA (with the opposite strand read as G), are methylated. Similarly, a symmetrically unmethylated site is represented by uC+uG. Our analysis shows that when iPGCs are differentiated from ESCs, there is a loss in symmetrical methylation and a 3-fold increase in

asymmetrical methylation (Figure 3-1J). Taken together, differentiation of iPGCs from ESCs results in a genome-wide reduction in cytosine methylation similar to what was previously reported for immature PGCs in phase 1 [11].

Tet proteins and 5hmC are found in PGCs in vivo and in vitro

Given the role for Tet mediated conversion of 5mC to 5hmC as an intermediate in locus-specific DNA demethylation [4], we were interested in examining the dynamic expression *Tets* at a single cell level in Oct4-GFP+ by sorting PGCs at e9.25, e10.25 and e11.5 (Figure 3-2A-D) as well as iPGCs sorted at day 6 of EB differentiation (Figure 3-2E). QRT-PCR and RNA-Seq has been used to evaluate Tet gene expression in Oct4-GFP+ PGCs [7, 14], however the heterogeneity between PGCs in sequential developmental ages has not been well defined. Single cell analysis revealed that *Tet1* is expressed as early as e9.25, and was detected in every *Dppa3*+ PGC examined until e11.5. In contrast, *Tet2* was heterogeneously expressed at e9.25 and e10.25 (51% and 57% of cells respectively). At e11.5, the number of *Tet2*+ PGCs increased substantially to be expressed in almost every *Dppa3*+ PGC together with *Tet1*. Unlike *Tet1* and *Tet2*, *Tet3* was expressed in rare *Dppa3*+ PGCs at e11.5. The cytidine deaminase *Aid* was negative suggesting that Aid does not act during this period. Comparably, by analyzing iPGCs, we discovered that 40/40 iPGCs expressed *Tet1*, and almost every single iPGC (39/40) also expressed *Tet2* (Figure 3-2E). Similar to PGCs from the embryo, we did not find *Aid* expression in iPGCs, and *Tet3* was rarely expressed (Figure 3-2E).

Given the expression of at least 1 or 2 *Tet* genes in PGCs, we analyzed 5hmC by immunostaining (Figure 3-2F-G). We used the commercially available 5hmC antibody that was previously confirmed as specifically recognizing 5hmC and not C or 5mC [27]. To confirm

specificity we transiently transfected 293T cells with a plasmid encoding the Tet1 catalytic domain (Tet1-CD) and show specific signal in transfected cells only (Figure 3-S2A). Next, using this antibody, we show that 5hmC is present in e10.5 SSEA1+ PGCs (arrow heads) at levels similar to somatic cells (Figure 3-2F). Expression of 5hmC was mostly uniform through the PGC nucleus. However, by e13.5, 5hmC exhibits a characteristic punctate pattern that overlaps with DAPI positive pericentromeric heterochromatin (Figure 3-S2B). Immunohistochemistry of sorted iPGCs and control SSEA1+ undifferentiated V6.5 ESCs revealed 5hmC enrichment in both cell types (Figure 3-2G) [28].

To quantify global levels of 5hmC and 5mC we used combined liquid chromatography electro-spray ionization tandem mass spectrometry with multiple reaction monitoring (LC-ESI-MS/MS-MRM) (Figure 3-2H and I). Using this technique 5.1% of total cytosines in V6.5 ESCs were methylated (Figure 3-2H) consistent with our previous report [29]. Next we found that the levels of 5mC in iPGCs were significantly reduced to on average 2.5%, while 5mC levels in somatic cells from the same EB was 4.9%. Furthermore, V6.5 ESCs and iPGCs had a similar 5hmC/5mC ratio where 5hmC is approximately 50-fold lower in abundance than 5mC (Figure 3-2I). In contrast, analysis of somatic cells from the EB at day 6 revealed a significant reduction in 5hmC relative to ESCs as previously reported [30, 31]. Taken together, differentiation of iPGCs from ESCs is not associated with a significant reduction in 5hmC compared to somatic cells, and suggests that iPGCs uniquely regulate the 5hmC modification during differentiation.

Tet1 and Tet2 do not regulate genome-wide demethylation in iPGCs

Given that 5hmC is found on less than 1% of cytosines in undifferentiated ESCs and iPGCs, and that *Tet1* and *Tet2* are co-expressed, we considered two alternate hypotheses for the role

of Tet1 and Tet2 in phase 1 genome-wide DNA demethylation. First, we could hypothesize that 5mC oxidation by *Tet1* and *Tet2* is required for genome-wide DNA demethylation in iPGCs, and this is initiated extensively in immature PGCs during EB formation. Therefore, the small measureable amounts of 5hmC in iPGCs at day six would underestimate the total 5mC to 5hmC conversion that occurred. A second hypothesis could be that 5hmC plays no role in phase 1 DNA demethylation in iPGCs.

To address these possibilities we designed an experiment to generate iPGCs from *Tet2-/-* ESCs transduced with a lentivirally (LV)-delivered shRNA against *Tet1* (shTet1 LV). We used this approach because iPGCs (and many endogenous PGCs) co-express *Tet1* and *Tet2* (Figure 3-2E), and depletion of *Tet1* and *Tet2* singly in ESCs results in only mild to moderate changes in 5hmC [30]. We found no statistically significant difference in the percentage of iPGCs differentiated from *Tet2* -/- ESCs transduced with a control LV (*Tet2-/-*; Control LV), relative to *Tet2* -/- ESCs transduced with the *Tet1* shRNA LV (Tet2-/-; shTet1 LV) (Figure 3-3A). To determine knockdown in the sorted iPGCs, we used FACS to isolate iPGCs and examined *Tet* gene expression by real-time PCR (Figure 3-3B). As a positive control for *Tet2* we sorted iPGCs differentiated from V6.5 ESCs transduced with control LV (WT; Control LV). *Tet1* RNA was successfully depleted with shTet1 LV in iPGCs up to 80% (Figure 3-3B). *Tet2* was undetectable in *Tet2-/-* ESCs, and *Tet3* levels were very low relative to Tet1, and were unchanged in the context of a *Tet2* deletion with or without a *Tet1* knockdown (Figure 3-3B,C). Therefore, modulating Tet gene expression in iPGCs does not result in compensatory expression of other Tets, similarly to what was reported in undifferentiated ESCs [30, 32]. We also analyzed PGC-expressed genes *Blimp1* (determinant of PGC fate) *Dppa3, Prdm14* (determinant of PGC fate) and *Dnd1*, and found that manipulation of *Tet1* and *Tet2* had no significant effect on expression of these genes relative to control (Figure 3-3D).

Finally, we asked if *Tet1* and *Tet2* regulate genome-wide DNA demethylation with iPGC differentiation. If we accept the hypothesis that Tet1 and/or Tet2 are required for regulating genome wide DNA demethylation, we would anticipate that cytosine methylation should significantly increase in Tet2-/-; shTet1 LV iPGCs compared to Tet2-/-; Control LV. To reject the hypothesis and find that Tet1 and Tet2 have no major role in genome-wide iPGC demethylation, we would expect that Tet2-/-: shTet1 LV iPGCs would have DNA methylation levels equivalent to wild type (Figure 3-1C). Using LC-ESI-MS/MS-MRM we found that 5hmC is no longer detectable in Tet2-/-; shTet1 iPGCs relative to Tet2-/-; control LV (Figure 3-3E). We also found that at the start of differentiation (day 6 post transduction) the Tet2-/-; shTet1 LV ESC samples had undetectable levels of 5hmC and no change in 5mC levels compared to Tet2-/-; control (Figure 3-S2C,D). Finally, analysis of 5mC by mass spectrometry revealed no increase in methylation in iPGCs with depleted 5hmC (Figure 3-3F). Thus we conclude that *Tet1* and *Tet2* do not regulate global DNA demethylation.

Using an alternate approach, whole genome BS-Seq in biological triplicate, we identified 6,866,888 CG dinucleotides that were represented in all six libraries and in agreement with the LC-ESI-MS/MS-MRM assay we show no significant change in the percentage of CG methylation between Tet2-/-; shTet1 LV iPGCs compared to Tet2-/-; Control LV (Figure 3-3G). Combined, our data reject the first hypothesis that *Tet1* and *Tet2* regulate genome wide DNA demethylation during iPGC differentiation. However, the BS-Seq results reveal a small (~4%) increase in methylation in the Tet1/2 mutant iPGCs, which suggests locus-specific effects. Indeed, we found that a knockdown of *Tet1* in a *Tet2* null background had a local effect on promoter and gene body methylation in iPGCs when compared to *Tet2-/-* iPGCs transduced

with a control LV. Importantly, the most significant directional change observed was CG hypermethylation in Tet2-/-; shTet1 LV iPGCs compared to Tet2-/-; Control LV reference iPGCs (Figure 3-3H and 3I). Some notable hypermethylated promoters included the genome defense genes *Tudor domain containing protein* (*Tdrd*) *Tdrd5, Piwi like 4 (Piwil4),* which are required later in germ cell development to repress transposons, and *Tdrd7, Spag8* and *Pramel1,* which are expressed in adult testis (Figure 3-3J). Curiously, gene body hypermethylation was discovered on germ cell expressed genes *Nanos3, Blimp1* and *Dppa2* and imprinted genes *Peg3* and *Snrpn* (Figure 3-3K). However, the increase in gene body methylation at *Blimp1* did not alter expression (Figure 3-3D). We noted that olfactory receptor (*Olfr*) RNA and microRNAs were highly represented in both promoter and gene body classifications, and shown here is an Olfr (*Olfr1226*) and miRNA (*mir133b*) that exhibited a hypomethylation in the Tet1/Tet2 mutant iPGCs (Figure 3-3J,K). Together this leads to a model where phase I PGC demethylation involving the bulk removal of DNA methylation is Tet1 and Tet2 independent.

DISCUSSION

Our genome-wide analysis of cytosine methylation using BS-Seq of iPGCs revealed a statistically significant and reproducible genome-wide depletion of cytosine methylation similar to what was recently reported by [11]. Specifically, the iPGC model reported here captures reorganization of cytosine methylation at meiotic CGIs relative to undifferentiated ESCs, and a heterogeneous population of PGCs undergoing global phase 1 DNA demethylation and the initiation of some locus specific Tet-dependent DNA demethylation in phase 2. Although Tet1 and Tet2 were recently reported to regulate locus-specific demethylation in PGCs[4, 14], it was unknown whether the initial global depletion of DNA methylation required a 5hmC intermediate. Here, we show that Tet1, Tet2 and 5hmC are dispensable for the initial global depletion of 5mC

from the PGC genome. Instead, Tet depletion induces promoter and gene body hypermethylation consistent with 5hmC as having a locus-specific role in DNA demethylation in PGCs [33].

Our genome-wide analysis reveals approximately 1,000 promoters and gene bodies that are differentially methylated in Tet1/2 mutant iPGCs with the vast majority exhibiting significant hypermethylation. Our data set revealed that *Dazl*, *Sycp3* and *Mael* promoter methylation levels were all increased in the Tet1/2 mutants as previously reported [14, 33]. However these changes did not reach statistical significance in our study perhaps because the iPGCs here are younger than e10.5. One interesting observation was the discovery of new Tet-regulated germ cell-expressed genes including *Tdrd5* and *Piwil4 (*also called *Miwi2)* and *Tdrd7* that function in the male germ line after e13.5 [34-36]. This suggests that 5hmC may prepare the germ line epigenome for future functional events unique to this lineage.

If 5mC oxidation and deamination[10] are not responsible for driving genome-wide depletion in DNA demethylation in PGCs, what could be the mechanism? One possibility is that the global demethylation is linked to abnormalities in replication-coupled methylation inheritance prior to e9.5. This could be caused by mis-localization, inactivation or repression of *DNA methyltransferase 1* (*Dnmt1*) [19] or its co-factor *Uhrf1* during DNA synthesis [37]. Hairpin BS sequencing [38], which detects methylation on complementary strands of DNA is one way to address this in the ESC to iPGC model.

In conclusion, we propose that differentiation of iPGCs represents a heterogeneous population of immature PGCs in the process of undergoing global genome-wide depletion of methylation during phase 1 and the beginning of Tet-dependent demethylation in phase 2. Our studies demonstrate that Tet1 and Tet2 do not regulate initial genome-wide depletion of 5mC (Figure 3-1A), and instead clarifies the model to demonstrate that Tet1 and Tet2 function to regulate locus-specific methylation during PGC development.

MATERIALS AND METHODS

Embryonic stem cell culture and iPGC Differentiation

Mouse ESC maintenance, differentiation, and isolation of iPGCs were performed as previously described [24]. For the Tet studies, lentiviruses were modified from [39] to carry shRNA directed against Tet1 mRNA harboring hygromycin resistance. Cells were transduced with indicated VSV-G pseudotyped virus at MOI 1 and selected in 200ug/mL hygromycin on inactivated DR4 MEFs. EB differentiation was performed with hygromycin in the media for six days prior to FACS.

Mouse Studies

Oct4-gfp embryos were used to isolate PGCs by FACS as previously described [24]. All research protocols were approved by the Animal Care Use Committee at UCLA.

Single Cell PCR

Single cell analysis was performed using the Fluidigm BioMark microfluidics PCR system as previously described and involved an 18-cycle Specific Target Amplification reaction [24]. Heatmaps of Ct values were generated with Fluidigm PCR data analysis software.

Whole Genome Bisulfite Sequencing

BS-Seq libraries were prepared as previously described [40]. Libraries were all single end reads containing pre-methylated illumina adaptors. Libraries were sequenced on either a Genome Analyzer II or HiSeq2000 and the percent methylation did not change between machines. Mapping was performed using BS Seeker where methylation levels for each cytosine were determined by measuring the ratio of Cs to Cs plus Ts that align to each genomic cytosine [26]. Read lengths ranged from 50 to 80 after trimming 20 bases from 3' end. Multiple reads mapped to the same location were considered only once. For deep sequencing of libraries mB557 and mB556 (Table 3-1) only cytosines with coverage $\geq$4 were included for the downstream analysis (478,482,437 cytosines). This set was used to acquire chromosome views, metaplots, Shannon Entropy, distribution plots and analysis of symmetry. To generate metaplot of genes, the transcription start and end sites of selected genes were fixed and the upstream, body and downstream regions were binned into windows. For each window, the average methylation level is calculated. Therefore a metagene plot summarized the average methylation level per window and is plotted from the upstream to downstream direction. Shannon entropy was calculated $-\sum_{1}^{n} m_i Log(m_i)$, where m_i is the methylation levels estimated at the i-th CG site. In the comparison we took the average of Shannon entropy (i.e., divided by n). To map differentially methylated sites (DMS) CG sites were selected using the criteria of delta methylation levels $\geq$30% and subject to two-way binomial tests. This yielded 11, 994, 107 CG sites for analysis with a FDR of $\leq$5%. In the scatter plots, the points show Δ methylation level between the two samples, versus the changes of expression (log2 microarray 1- log2 microarray 2). Distribution plots of methylation levels were calculated at each cytosine as the ratio of Cs to Cs plus Ts that align to each genomic cytosine. The methylation levels of all cytosines with coverage $\geq$4x were plotted as a histogram. To calculate symmetry, the methylation status of these cytosines is based on FDR=1% and sequencing error 1% [41]. As a result, the methylation status (methylated or not methylated) of all eligible cytosines were determined by counting among CpG pairs the

numbers of both methylated (mC+mG), one methylated and one unmethylated (mC+uG), and both unmethylated (uC+uG) in mBS48 and mBS49. For analysis of Tet2-/-;LV and Tet2-/-;Tet1sh LV samples n=3 libraries were prepared in biological triplicate in samples that exhibited a knockdown of Tet1 which was ≥69% of the Tet2-/-;Control sample. This analysis yielded 6,866,888 CG dinucleotides that were represented in all six libraries and used to calculate %CG methylation. T-Tests were used to calculate significance between two groups. All data has been downloaded to Gene expression Omnibus (GEO).

LC-ESI MS/MS-MRM

DNA was extracted with the Zymo gDNA Mini Prep kit from freshly sorted cell populations by FACS, and LC-ESI-MS/MS-MRM was used to determine the proportional content of 5mC or 5hmC relative to total cytosine levels, according to [29].

Immunofluorescence

Staining of paraffin-embedded tissue was performed with conventional protocols with some exceptions. For 5mC and 5hmC staining, tissue was first stained for other markers, post-fixed, and denatured for 10 minutes in 4N HCl prior to overnight incubation with appropriate antibodies. The following antibodies were used: Oct4 (Santa Cruz), E-cadherin (BD), SSEA1 (DSHB), 5mC (Aviva), and 5hmC (Active Motif). Fluorescent visualization was performed using isotype-specific secondary antibodies conjugated to either FITC or Alexa-Fluor 594 (Jackson ImmunoResearch). All images were acquired on a Zeiss LSM 780 confocal microscope.

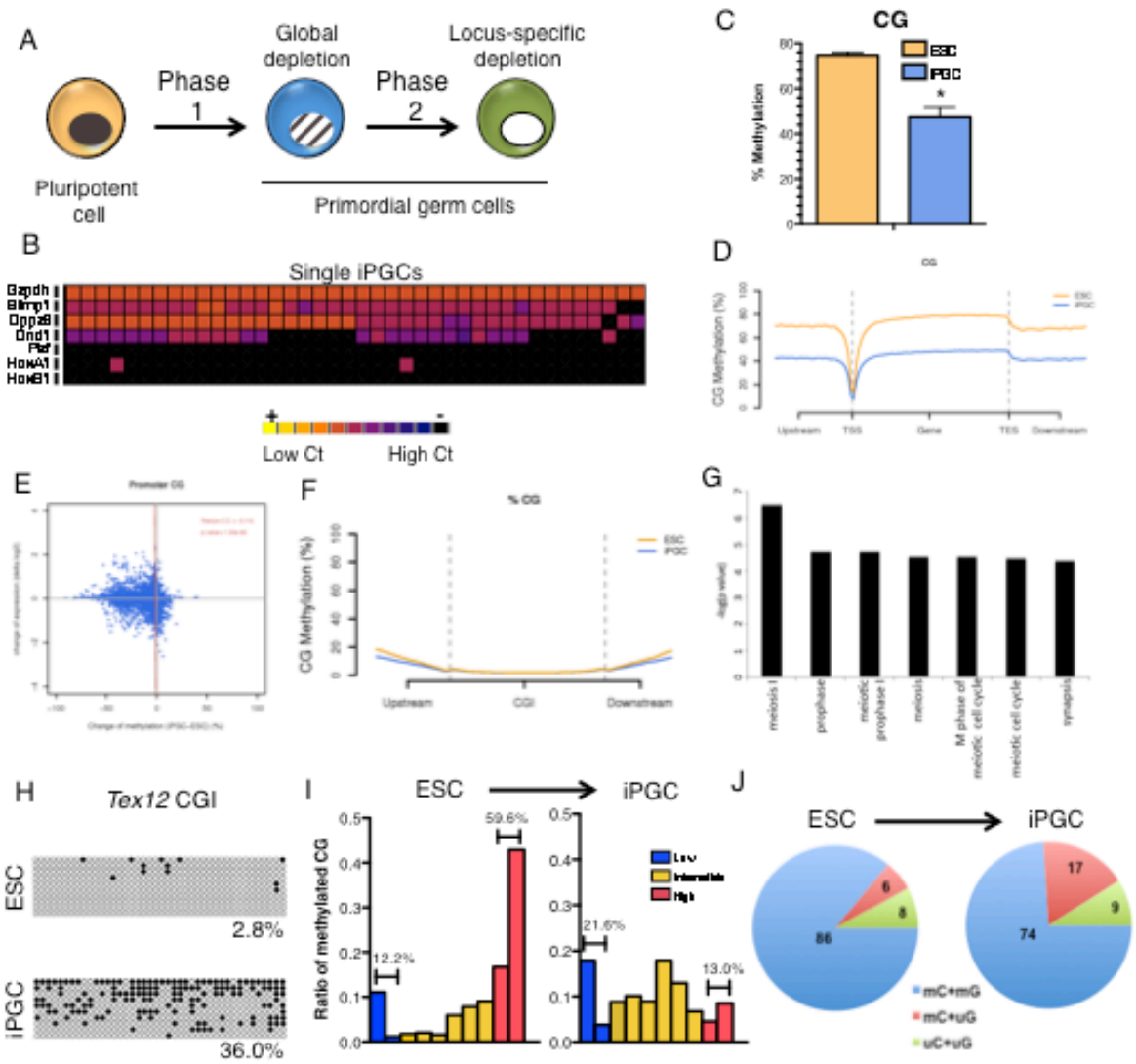

Figure 3-1. Generation of PGCs from ESCs results in a significant decrease in CG methylation.

A: Two-phase model of PGC demethylation. PGCs are specified from pluripotent cells (yellow), and initially contain high levels of 5mC (black nucleus). In phase 1, PGCs younger than e9.5 (blue) undergo global DNA demethylation [11]. In phase 2 PGCs undergo locus-specific demethylation (white nucleus). Tet1 & Tet2 regulate locus-specific demethylation in phase 2 [4]. The role of Tet1 and Tet2 in the global demethylation is unknown. B: Single cell analysis of sorted iPGCs. C: Quantification of cytosine methylation in CG sequence context by BS-Seq. Shown is mean ±SD (n=3). D: Metaplot analysis of CG methylation at RefSeq genes. E: Pearson analysis of differentially methylated sites (DMS) with gene expression. F: Metaplot of CG methylation at CG islands (CGIs). G: Gene ontology analysis of 100 CGIs with significantly higher methylation levels in iPGCs. H: Bisulfite-PCR of *Tex12* CGI. Black circles = methylated cytosines; white circles = unmethylated cytosines. I: Distribution of cytosine methylation in ESCs and iPGCs. Binned bars representing percent methylation are graphed along x-axis. J: Frequency of methylation symmetry in CG sequence contexts in iPGCs after differentiation from

ESCs. * indicates $p<0.05$. Figure 3-3-1. Generation of PGCs from ESCs results in a significant decrease in the amount of CG methylation.

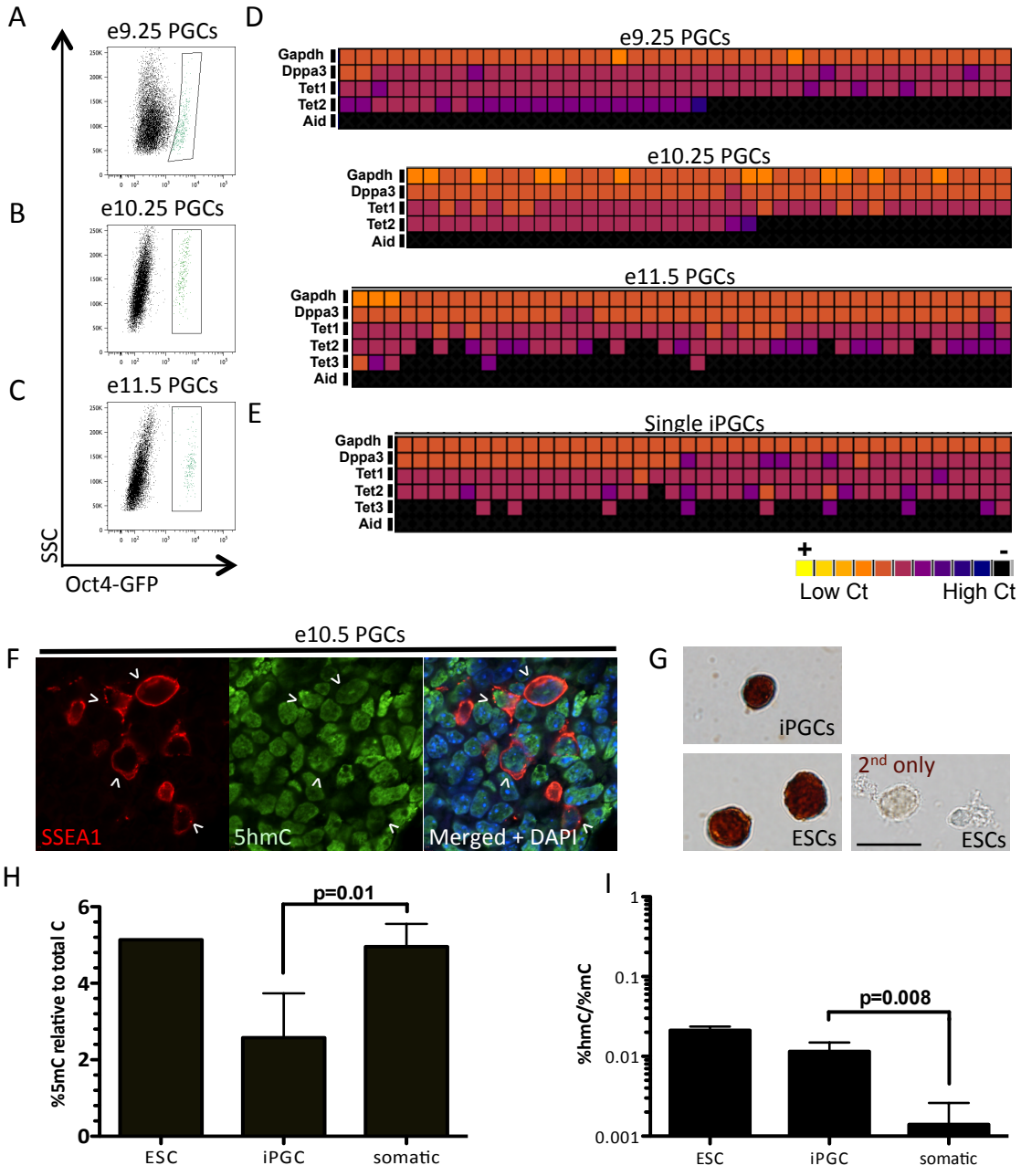

Figure 3-2. Tet genes and 5-hydroxymehtylcytosine are present in PGCs and iPGCs.

A-C: FACS plots of GFP+ PGCs (green) sorted from the mouse embryo at time points indicated. D: Single cell analysis of e9.25 (A), e10.25 (B), and e11.5 (C) GFP+ PGCs for *Tet1, Tet2, Tet3* and *Aid.* E: Single cell analysis of day 6 iPGCs. F: Immunofluorescence of e10.5 PGCs for SSEA1 (red) and 5hmC (green). Arrowheads depict SSEA1+ PGCs with 5hmC content. G: Immunohistochemistry for 5hmC in sorted day 6 iPGCs and undifferentiated ESCs. Control involves omitting primary antibody. H and I: Mass spectrometry analysis of ESCs, iPGCs, and somatic EB cells for 5mC (H) and 5hmC (I) content.

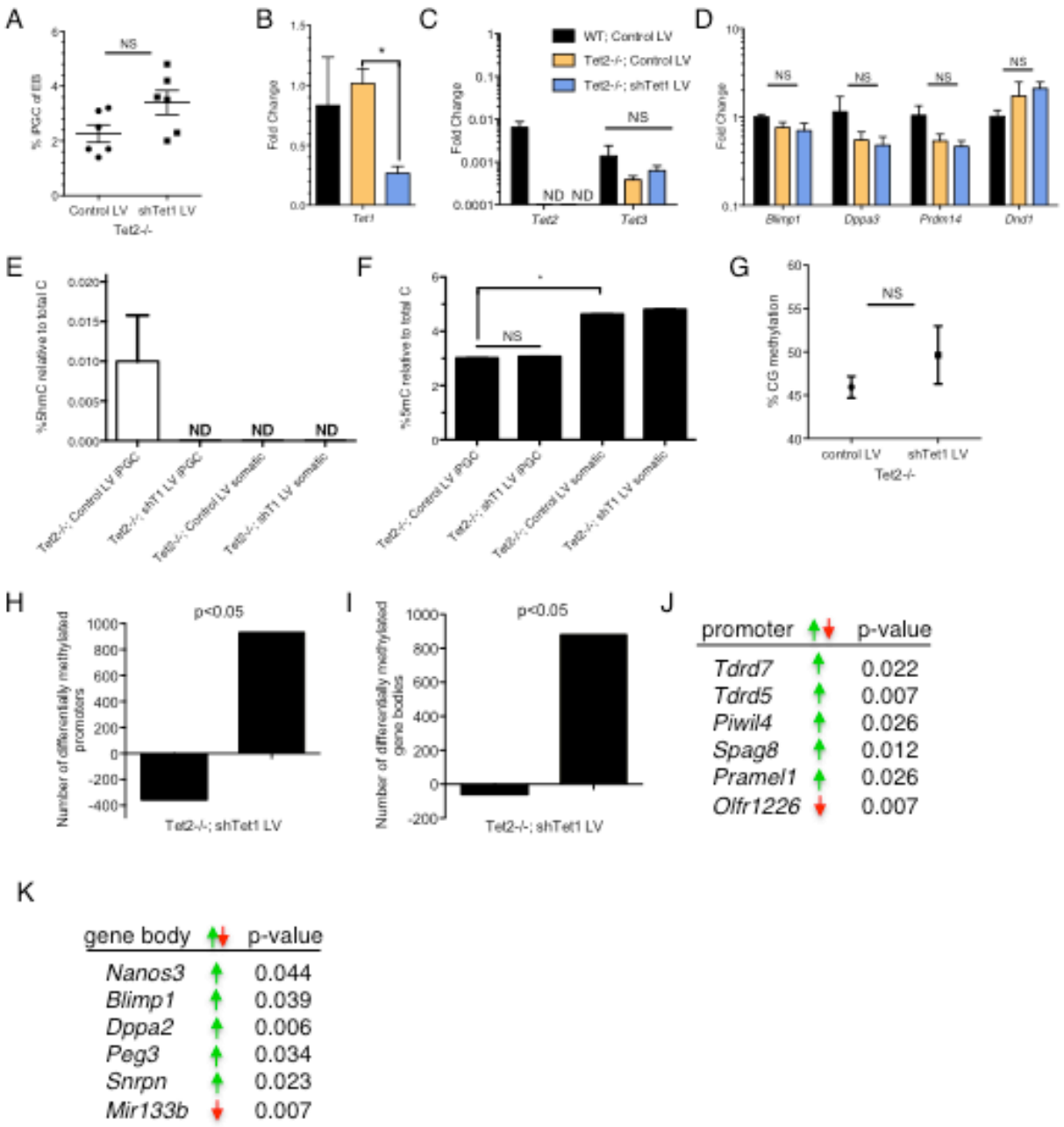

Figure 3-3. Tet1 and Tet2 do not regulate phase 1 global DNA demethylation in iPGCs.

A: Yield of iPGC differentiations in Tet2-/- cells transduced with a control lentivirus (LV) and a LV containing a shRNA against *Tet1* (shTet1 LV*)* (mean ± SEM of n=6). B-D: Real-time PCR (mean ± SEM of n=3) E,F: LC-ESI-MS/MS-MRM measurements for 5hmC (E) and 5mC (F) in iPGCs and somatic cells (mean ± SD of n=3). G: Genome-wide CG methylation levels by BS-Seq calculated from paired experiments (mean ± SEM of n=3). H,I: Number of promoters (H) and gene bodies (I) that exhibit a statistically significant (p<0.05) decrease (negative) or increase (positive) in CG methylation in Tet2-/-; shTet1 LV PGCs. J-K: Promoters (J) and gene bodies (K) with differential methylation in Tet2-/-;shTet1 iPGCs (green arrow = increase in methylation, red arrow = decrease in methylation). (NS=not significant, ND=not detected) * indicates p<0.05.

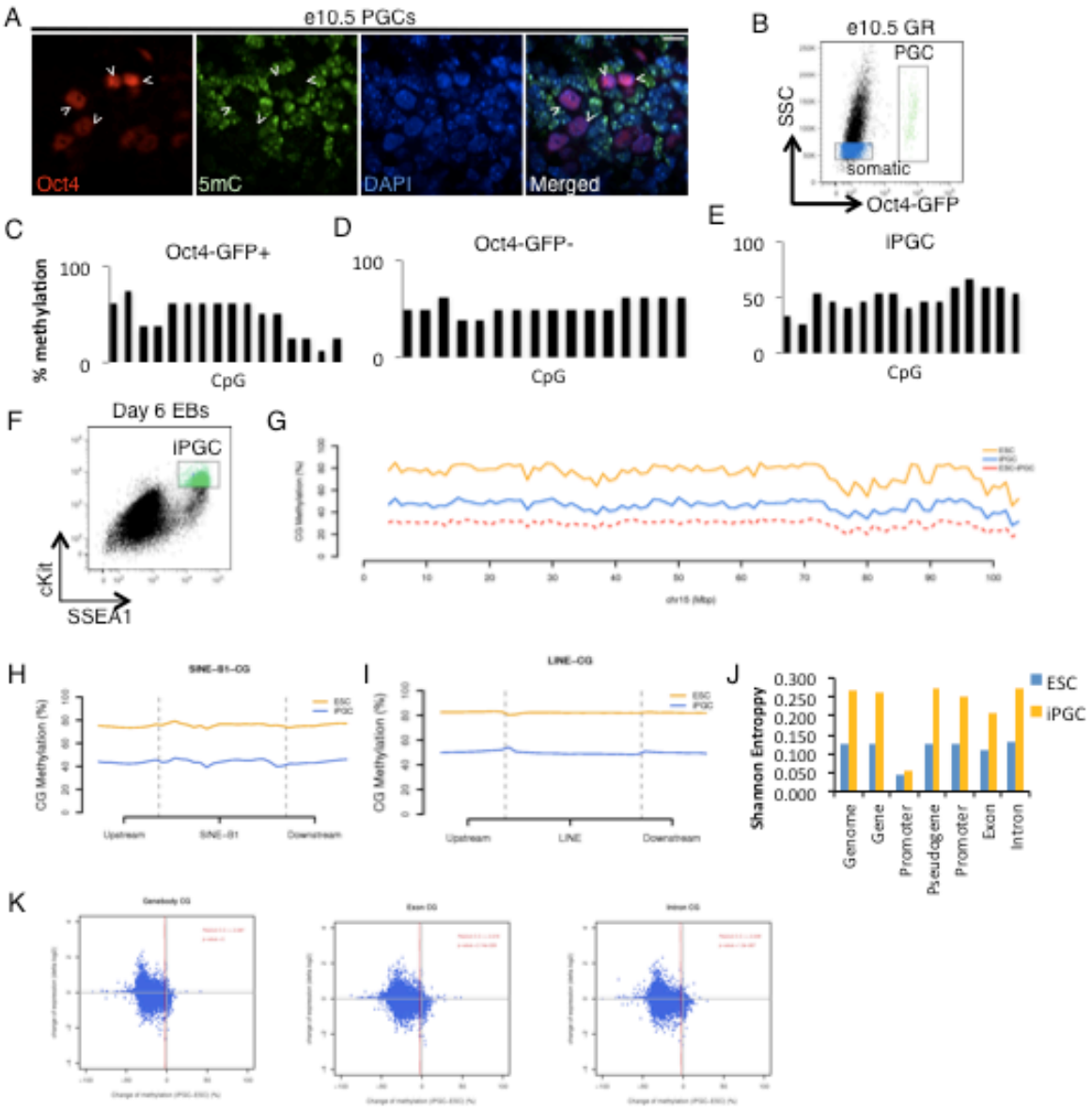

Figure 3-S1. DNA methylation in ESCs and iPGCs.

A: Immunofluorescence of e10.5 genital ridges for Oct4 (red) and 5mC (green). Arrowheads denote Oct4+ PGCs. B: Flow cytometry of e10.5 *Oct4-gfp* genital ridges (GR). GFP+ cells (green) are PGCs. GFP- somatic cells (blue). C,D,E: Percent methylation at *Snrpn* ICC in GFP+ PGCs (C), GFP- somatic cells from e10.5 embryos (D) and iPGCs (E). (F): Flow cytometry of V6.5 EBs at Day 6 of differentiation, showing gating strategy for SSEA1+/cKit^{bright} iPGCs (green). G: Metaplot of methylation across murine chromosome 15. Methylation percentage (y-axis) is graphed along the coordinate distance of the chromosome in megabases (x-axis). H,I: Methylation of SINE B1 (H) and LINE (I) transposable elements. J: Shannon Entropy analysis of ESCs and iPGCs at genomic regions indicated. K: Pearson correlations of change of methylation between iPGCs and ESCs against gene expression or repression. Correlations are performed on CG methylation found in gene body, exon, and intronic regions. Scale bar=5um.

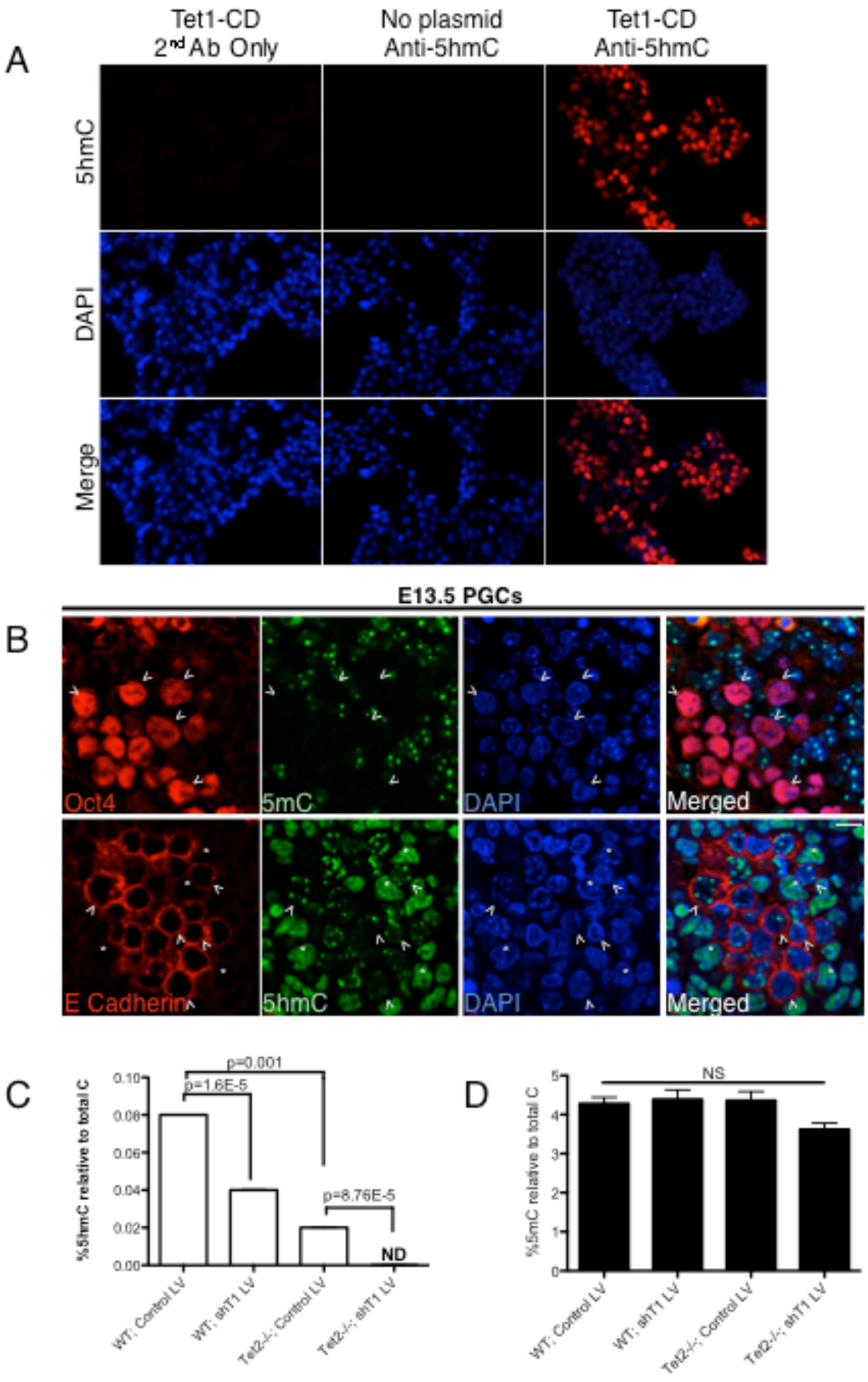

Figure 3-S2. 5hmC analsyis of ESCs and IPGCs.

A: HEK293 cells, which do not have detectable 5hmC, were transfected with Tet1-CD overexpression construct. Staining for 5hmC (red) was performed as indicated. 5hmC signal was only detectable by fluorescent microscopy with transfection of Tet1-CD. B: 5hmC (green) staining of Oct4 and E-Cadherin (red) positive germ cells shows a punctate expression at e13.5. Scale bar=5um. C: 5hmC in undifferentiated ESCs. D: 5mC in undifferentiated ESCs. Measurements are displayed as mean $\pm$ SD n=3. Significance is shown on graph.

Sample	Library	CG	CHG	CHH	CA	CC	CT
iPGC	mBS44	52.19%	0.70%	0.47%	0.95%	0.16%	0.27%
ES	mBS47	74.97%	1.43%	0.97%	2.15%	0.19%	0.49%
iPGC	mBS48	44.74%	0.56%	0.40%	0.76%	0.17%	0.27%
ES	mBS49	73.39%	0.77%	0.52%	1.16%	0.14%	0.28%
iPGC	mBS56	45.16%	0.42%	0.31%	0.55%	0.15%	0.22%
ES	mBS57	75.92%	0.77%	0.55%	1.12%	0.16%	0.31%

Table 3-1. BS-Seq of pair-wise differentiations of iPGCs and starting undifferentiated ESCs.

Methylation was calculated in various genomic contexts. Black: Differentiations performed with R26-GFP ESCs. Blue: Experiments using V6.5 ESCs and iPGCs. Deep sequencing analysis was performed on libraries mBS48 and mBS49.

REFERENCES

1. Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8689-8694 (2010).

2. Gkountela, S. *et al.* The ontogeny of cKIT(+) human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat Cell Biol* **15**, 113-122 (2012).

3. Guibert, S., Forne, T. & Weber, M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome research* **22**, 633-641 (2012).

4. Hackett, J.A. *et al.* Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. *Science* (2012).

5. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-881 (2008).

6. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23 (2002).

7. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78-82 (2010).

8. Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371-382 (1987).

9. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

10. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101-1105 (2010).

11. Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular cell* (2012).

12. Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**, 440-458 (2005).

13. Smith, Z.D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-344 (2012).

14. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* (2012).

15. Lawson, K.A. & Hage, W.J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Foundation symposium* **182**, 68-84; discussion 84-91 (1994).

16. Ohinata, Y. *et al.* Germline recruitment in mice: a genetic program for epigenetic reprogramming. *Ernst Schering Research Foundation workshop*, 143-174 (2006).

17. Tam, P.P. & Zhou, S.X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* **178**, 124-132 (1996).

18. Ying, Y., Qi, X. & Zhao, G.Q. Induction of primordial germ cells from murine epiblasts by synergistic action of BMP4 and BMP8B signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7858-7862 (2001).

19. Lei, H. *et al.* De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195-3205 (1996).

20. Kafri, T. *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev* **6**, 705-714 (1992).

21. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nature reviews. Genetics* **2**, 21-32 (2001).

22. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**, 88-93 (2003).

23. Lees-Murdock, D.J., De Felici, M. & Walsh, C.P. Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics* **82**, 230-237 (2003).

24. Vincent, J.J. *et al.* Single cell analysis facilitates staging of Blimp1-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS One* **6**, e28960 (2011).

25. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* **5**, e8888 (2010).

26. Chen, P.Y., Cokus, S.J. & Pellegrini, M. BS Seeker: precise mapping for bisulfite sequencing. *BMC bioinformatics* **11**, 203 (2010).

27. Iqbal, K., Jin, S.G., Pfeifer, G.P. & Szabo, P.E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3642-3647 (2011).

28. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398-402 (2011).

29. Le, T., Kim, K.P., Fan, G. & Faull, K.F. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Analytical biochemistry* **412**, 203-209 (2011).

30. Koh, K.P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell* **8**, 200-213 (2011).

31. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935 (2009).

32. Dawlaty, M.M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell stem cell* **9**, 166-175 (2011).

33. Hackett, J.A., Zylicz, J.J. & Surani, M.A. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends in genetics : TIG* **28**, 164-174 (2012).

34. Carmell, M.A. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**, 503-514 (2007).

35. Tanaka, T. *et al.* Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10579-10584 (2011).

36. Yabuta, Y. *et al.* TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. *The Journal of cell biology* **192**, 781-795 (2011).

37. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764 (2007).

38. Laird, C.D. *et al.* Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 204-209 (2004).

39. Ambartsumyan, G. *et al.* Centromere protein A dynamics in human pluripotent stem cell self-renewal, differentiation and DNA damage. *Hum Mol Genet* **19**, 3970-3982 (2010).

40. Feng, S., Rubbi, L., Jacobsen, S.E. & Pellegrini, M. Determining DNA methylation profiles using sequencing. *Methods in molecular biology (Clifton, N.J.)* **733**, 223-238 (2011).

41. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).

CHAPTER 4

THE ONTOGENY OF cKIT+ HUMAN PRIMORDIAL GERM CELLS: A RESOURCE FOR HUMAN GERM LINE REPROGRAMMING, IMPRINT ERASURE, AND *IN VITRO* DIFFERENTIATION

Preface

With the development of technologies capable of interrogating the molecular identity of cell lineages at the single-cell level, many groups have sought to understand transcriptional dynamics of individual cells, rather than populations of cells, which may be heterogeneous in terms of their transcription and function. Single cell analysis is also highly amenable to study of early embryos, which generally have limited material for analysis. Techniques and analytical platforms, ranging from quantitative PCR [1, 2], to global transcriptome profiling by arrays[3, 4] and by next-generation sequencing [5-7] have been developed, and have provided new insights to transcriptional changes that accompany embryonic development.

In mouse, single cell analysis has provided insight into pre-implantation development. Analysis of single epiblast cells, PGCs, and embryonic somatic cells identified differentially expressed genes between these lineages, including the receptor tyrosine kinase *cKit,* which is restricted to PGCs and ESCs, but not epiblast [2]. Whole-genome transcriptional analysis by RNA-seq of single oocytes and blastomeres uncovered thousands of genes expressed in these cell types which were previously unknown by conventional array approaches[8]. Comparison of ICM blastomeres, newly derived ESC outgrowths, and established ESC lines identified an ESC-specific transcriptional signature [5]. Gene ontology analysis of the transcriptomes of ICM blastomeres, ICM outgrowths and ESC lines identified differentially expressed genes that increase expression upon ESC line establishment, including epigenetic regulators (such as DNA methyltransferases and histone deacetylases), as well as differential expression of metabolic genes [5]. Single cell technology has also been applied to early human development. Analysis of early human embryos from zygote through blastocyst stage has identified temporal-specific gene expression signatures over this developmental period [9]. Two-cell human embryos with functional defects in blastocyst formation *in vitro* were found to display a transcriptional

signature associated with arrested development, characterized by diminished levels of genes associated with cytokinesis, pluripotency, and RNA processing [9].

In the case of PGCs, many groups have used single cell technology to interrogate cell identity at the transcriptional level by PCR [1, 10-13]. Single cell analysis identified a core signature of miRNAs expressed by purified PGCs starting at e9.5, and uncovered a role for miRNA biogenesis in PGC development and germ cell competency for spermatogenesis [14]. Combining mouse genetics with single cell technology, one study used single cell analysis to define a comprehensive transcriptional network required to sustain PGC formation through the action of *Blimp1* [4].

Single cell approaches have also been applied to pluripotent-based differentiation strategies to compare developmental progression in a dish to *in vivo* development and have been indispensible in understanding heterogeneity of the cell types that can be differentiated in vitro, which are ultimately lost by conventional gene analysis of pooled sample [1, 11, 15]. The importance of single cell analysis of PGC formation was particularly evident from the studies presented in Chapter 2, which revealed that single ESCs cells individually express stochastic combinations of PGC-associated genes—rather than a population of cells expressing PGC transcripts at low levels [1]. This increased resolution has facilitated examination of developmental progression in multiple early developmental lineages, and is highly compatible with small sample sized that can exist in early embryos. Therefore, the development of single cell technology has provided a new level of resolution in understanding complex biological variability and temporal regulation of specific RNAs, and has been instructional in the development of *in vitro* differentiation strategies.

In an effort to understand human germ cell development, we sought to apply a single-cell analysis to map developmental progression of human gonadal PGCs. In humans, PGC establishment largely occurs before a woman is pregnant. Combined with ethical considerations, studies of primary purified human fetal germ cells have been limited. To begin to map major developmental and epigenetic landmarks in human germ line development, we examined the gonads of human fetuses from terminations between 6-20 weeks of development, isolated PGCs by expression of the cell surface receptor *cKIT*, and identified key transcriptional and epigenetic events that occur during human fetal life. We also used quantitative PCR of single cells to evaluate multiple differentiation strategies of human ESCs to generate iPGCs *in vitro.*

REFERENCES

1. Vincent, J.J. *et al.* Single cell analysis facilitates staging of Blimp1-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS One* 6, e28960 (2011).

2. Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y. & Saitou, M. Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. *Biology of reproduction* 75, 705-716 (2006).

3. Kurimoto, K. & Saitou, M. A global single-cell cDNA amplification method for quantitative microarray analysis. *Methods in molecular biology (Clifton, N.J.)* 687, 91-111 (2011).

4. Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* 22, 1617-1635 (2008).

5. Tang, F. *et al.* Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell stem cell* 6, 468-478 (2010).

6. Tang, F. *et al.* Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS One* 6, e21208 (2011).

7. Tang, F., Lao, K. & Surani, M.A. Development and applications of single-cell transcriptome analysis. *Nature methods* 8, S6-11 (2011).

8. Lao, K.Q. *et al.* mRNA-sequencing whole transcriptome analysis of a single cell on the SOLiD system. *Journal of biomolecular techniques : JBT* 20, 266-271 (2009).

9. Wong, C.C. *et al.* Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature biotechnology* 28, 1115-1121 (2010).

10. Chu, L.F., Surani, M.A., Jaenisch, R. & Zwaka, T.P. Blimp1 expression predicts embryonic stem cell development in vitro. *Curr Biol* 21, 1759-1765 (2011).

11. Haston, K.M., Tung, J.Y. & Reijo Pera, R.A. Dazl functions in maintenance of pluripotency and genetic and epigenetic programs of differentiation in mouse primordial germ cells in vivo and in vitro. *PLoS ONE* 4, e5654 (2009).

12. Hayashi, K. *et al.* MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS One* 3, e1738 (2008).

13. Nicholas, C.R., Haston, K.M., Grewall, A.K., Longacre, T.A. & Reijo Pera, R.A. Transplantation directs oocyte maturation from embryonic stem cells and provides a therapeutic strategy for female infertility. *Hum Mol Genet* 18, 4376-4389 (2009).

14. Hayashi, K. *et al.* MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis. *PLoS ONE* 3, e1738 (2008).

15. McKinney-Freeman, S. *et al.* The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell stem cell* 11, 701-714 (2012).

The ontogeny of cKIT^{+} human primordial germ cells: A resource for human germ line reprogramming, imprint erasure and *in vitro* differentiation

Sofia Gkountela[1,2], Ziwei Li[1,2], John J. Vincent[1,2,3], Kelvin X. Zhang[4], Angela Chen[5], Matteo Pellegrini[1] and Amander T. Clark[1,2,3,6]

1 Department of Molecular Cell and Developmental Biology, 2 Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, 3 Molecular Biology Institute, 4 Department of Biological Chemistry, Howard Hughes Medical Institute, 5 Obstetrics & Gynecology, David Geffen School of Medicine, 6 Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, California, 90095; United States of America

ABSTRACT

Generation of research quality, clinically relevant cell types *in vitro* from human pluripotent stem cells (hPSCs) requires detailed understanding of the equivalent human cell types. Here we analyzed 134 human embryonic and fetal samples from 6-20 developmental weeks and identified the stages in which cKIT^{+} primordial germ cells (PGCs), the precursors of gametes, undergo whole genome epigenetic reprogramming with global depletion of 5mC, H3K27me3, H2A.Z and the time where imprint erasure is initiated and 5hmC is present. Using five alternate *in vitro* differentiation strategies combined with single-cell microfluidic analysis and a *bona fide* human cKIT^{+} PGC signature, we show the stage of cKIT^{+} PGC formation in the first 16 days of differentiation. Taken together, our study creates a resource of human germ line ontogeny that is essential for future studies aimed at *in vitro* differentiation and unveiling mechanisms necessary to pass human DNA from one generation to the next.

INTRODUCTION

The foundation of human health at a cellular and molecular level is built upon accurate lineage differentiation during embryonic and fetal life. In recent years, a major barrier to study human development was overcome through the generation of human pluripotent stem cells (hPSCs), including human embryonic stem cells (hESCs) and human induced pluripotent stem (hIPS) cells which can be used to differentiate to embryonic and fetal cell types. However, a major caveat for using hPSCs as a surrogate model for human fetal development is the dearth of studies that provide accurate human-specific details to validate, guide and quality control differentiation *in vitro*.

All adult human cells are created from four major embryonic lineages, ectoderm, mesoderm, endoderm and the germ line. The first three lineages contribute a variety of cell types to multiple organs. In contrast, the germ line has one purpose that is to generate gametes, which function solely to pass DNA from one generation to the next. There is considerable interest in generating germ line from hPSCs cells, as they could serve as a potential stem cell-based intervention for infertility[1,2], or a model to understand the genetic-basis of human infertility[3]. However, before this can be achieved, the major landmarks of human germ line development during embryonic and fetal life must be characterized.

Human germ line development begins with the formation of primordial germ cells (PGCs) that express the tyrosine kinase receptor cKIT[4-10]. Very little is known about the developmental progression (ontogeny) of cKIT$^+$ PGCs, however based on the mouse model, it is clear that PGCs must undergo whole genome epigenetic reprogramming in order to remove cytosine methylation from imprinted genes and restore totipotency[11-14]. Given the fundamental role of

epigenetic reprogramming in the germ line, it is essential to characterize reprogramming in human PGCs since mouse and human genomes are separated by ~170 million years and diverse strategies may have evolved to execute it. Once the major molecular landmarks of human PGC reprogramming are known, we propose that this information will be critical to assessing PGC differentiation and reprogramming *in vitro,* or identifying bottlenecks that must be overcome to generate a functional germ line from hPSCs.

By evaluating 134 human embryonic and fetal gonadal samples from 6-20 developmental weeks, we provide the first comprehensive transcriptional and epigenetic roadmap of human cKIT^{+} PGCs in testes and ovaries and pinpoint the timing of major epigenetic events including whole genome reprogramming and initiation of imprint erasure. Using the endogenous human cKIT^{+} PGCs as a reference, we can now more accurately interpret the identity of the cKIT^{+} subpopulation of PGCs acquired with *in vitro* differentiation from hESCs. Our results clearly demonstrate that single cell analysis at both RNA and protein level is critical to defining PGC identity *in vitro,* and indisputably shows that established hESC lines are not equivalent to human PGCs.

RESULTS

cKIT positive PGCs undergo molecular progression with fetal development

Temporal and spatial expression of cKIT in fetal testes and ovaries from 7-19 weeks of development was evaluated by immunofluorescence together with the evolutionarily conserved germ cell marker VASA. All testes samples procured had characteristic seminiferous cords by histology indicating that sex determination had been initiated[15] (Supplementary Figure 4-S1).

We identified cKIT on the surface of all VASA positive cells in testes from 7-11 weeks and ovaries from 7-9.5 weeks (Figure 4-1a,b, and Supplementary Figure 4-S2a). However, from 12.5 weeks in testes, and 11 weeks in ovaries, cKIT and VASA protein expression becomes uncoupled, with only 10% of $cKIT^+$ cells co-expressing VASA (arrows in Figure 4-1a,b, quantified in Figure 4-1c,d and Supplementary Figure 4-S2b). Upon uncoupling, the ratio of single $cKIT^+$ to single $VASA^+$ cells was 1:1. We also evaluated SSEA1, and found that although PGCs are $SSEA1^+$ in fetal testes at the "common PGC progenitor stage" and after cKIT/VASA uncoupling, SSEA1 alone is not specific for the human germ line since it was also expressed on cKIT and VASA negative cells (not germ cells) (Supplementary Figure 4-S3a,b).

To assess the stem cell identity of $cKIT^+$ gonadal cells we examined the germ/stem cell enriched protein OCT4A using antibodies against the N-terminal region that discriminates OCT4A from the splice variant OCT4B[16,17] (Figure 4-1e,f). OCT4A localized to the nucleus of $cKIT^+$ cells from 7-10.5 weeks in testes and 6-8.5 weeks in ovaries (just prior to VASA repression). Similarly, expression of the pluripotency marker TRA-1-81 highly correlated with nuclear OCT4A in both sexes (Supplementary Figure 4-S3c,d). After this time, our data indicates that $OCT4A^+$ cells become a subpopulation of $cKIT^+$, and the majority of $cKIT^+$ PGCs localize OCT4A protein to the cytoplasm. In addition a number of $cKIT^+$ cells no longer express OCT4A (Figure 4-1g,h). At 17 weeks in fetal testes and from 16.5 weeks in fetal ovaries OCT4A is again identified in the nucleus of a large fraction of $cKIT^+$ cells (Figure 4-1g,h). Therefore, using cKIT, OCT4A and VASA expression, we propose a common PGC progenitor stage in humans that lasts to 11 weeks in testes and 9.5 weeks in ovaries. Thereafter two major populations are established in males and females, the $cKIT^+$ population that expresses OCT4A in most cells, and the single $VASA^+$ cells.

To isolate individual cKIT$^+$ cells, we performed fluorescence-activated cell sorting (FACS) of 49 testes and 42 ovaries from 8-20 developmental weeks (Table S1), using the gating strategy shown in Figures 2a,b. Applying this sorting strategy on a 15.5-week testis, in combination with quantitative (q) qRT-PCR, we verified that germ line identity was specifically enriched in the cKIT bright fraction compared to cKIT dim or single SSEA1 expressing cells (Supplementary Figure 4-S4a-c). We speculate that the cKIT dim gate is a heterogeneous mixture of PGCs and non-PGCs given that all germ line genes including *VASA* are reduced relative to the cKIT bright fraction. Therefore, to avoid potential contamination with gonadal somatic cells in down-stream applications, we excluded cKIT dim cells from all future FACS. Using the cKIT bright gate (which we call cKIT$^+$) we sorted an average of 2.83% cKIT$^+$ cells per testis at 8-11 weeks and 2.45% cKIT$^+$ cells per ovary at 8-9.5 weeks. Then at 11.1-20 weeks we sorted an average of 0.9% cKIT$^+$ cells from individual testes and 4.75% cKIT$^+$ cells per ovary at 9.6-16.5 weeks (Figure 4-2a,b). The absolute number of cKIT$^+$ cells sorted from an individual testis was as low as 150 for a 20-week sample, with the majority of samples yielding 2,500-3,000 cKIT$^+$ cells per testis. Similarly, fetal ovaries yielded on average 4,500-5,000 cells per ovary, ranging from as low as 276 cells for an 8-week ovary to 30,000 cells, for a 16.5-week ovary. This range in absolute numbers most likely reflects variability in sample quality (intact gonads verses fragments) and viability (which ranged from 14.4%-64.7%). However, the variability in the percentage of cKIT$^+$ cells in the ovary did not correlate with overall sample viability, and instead we speculate that this variability was due to the presence of small amounts of attached non-gonadal tissue that varied from sample to sample.

To determine the molecular identity of cKIT$^+$ PGCs, we performed single cell analysis with five PGC signature genes including *OCT4*, *BLIMP1, DAZL*, *VASA* and *NANOS3* using FACS, followed by microfluidic qRT-PCR at the common progenitor and cKIT/VASA uncoupled stage

(Figure 4-2c-g). We also confirmed expression of cKIT in individual cells (Figure 4-2f). Our single cell approach was first validated in HEK 293 cells (Supplementary Figure 4-S4e-h). At the common PGC progenitor stage, 14/16 cKIT$^+$ cells in the testis, and 13/20 cKIT$^+$ cells in the ovary coordinately expressed the five PGC signature genes (Figure 4-2c,d). However, in the ovary 7/20 cKIT$^+$ cells did not express *VASA* and/or *DAZL* at this stage and instead were *OCT4/BLIMP1* double positive (O/B) or *OCT4*/*BLIMP1*/*NANOS3* triple positive (O/B/N3). In testes, *NANOS2* expression was also evaluated and found in <20% of cKIT$^+$ cells in the common progenitor and this was maintained upon cKIT/VASA uncoupling in the cKIT$^+$ cell (Figure 4-2c,e). In the ovary during the uncoupled stage when OCT4A is either in the cytoplasm or no longer expressed, *NANOS3* mRNA is also no longer expressed in a fraction of cells, and these *NANOS3* negative cells correlated with no or low levels of *OCT4* and *BLIMP1* (Figure 4-2f). At 16.5 weeks, when OCT4A is again localized to the nucleus or not expressed, ovarian cKIT$^+$ cells become even more heterogeneous, most notably involving loss of *NANOS3*, *OCT4 and BLIMP1* mRNA in some cells, with *DAZL* and *VASA* being absent in others (Figure 4-2g). *SYCP3* was used to indicate meiotic potential, and was expressed in every cell at 14 and 16.5 weeks (Figure 4-2f,g). Furthermore, at 16.5 weeks *SYCP3* and *VASA* mRNA expression levels were significantly enriched in the *NANOS3* negative population (Supplementary Figure 4-S4d). Despite *SYCP3* mRNA expression in every cell at a single cell level, on the protein level, only VASA$^+$ cells are immunopositive for SYCP3 in the fetal ovary from 14 weeks and not cKIT$^+$ (Supplementary Figure 4-S3e), indicating that VASA$^+$ cells are the first to acquire meiotic potential.

Loss of 5mC from imprinted DMRs is locus specific and occurs weeks after global 5mC depletion

By immunofluorescence, 5-methyl cytosine (5mC) was below the level of detection at all stages of PGC development compared to somatic cells (open arrowheads on Figure 4-3a,b). To evaluate cytosine methylation at differentially methylated regions (DMRs) of imprinting control centers, we used Bisulfite Sequencing (BS) followed by PCR (BS-PCR) on cKIT$^+$ sorted PGCs (Figure 4-3c,d). We evaluated two paternally methylated DMRs, *H19* and *MEG3* and two maternally methylated DMRs, *PEG3* and *KCNQ1*. Primers were first verified using the BJ primary fibroblast cell line and H1 hESCs (Supplementary Figure 4-S5a). For the paternally methylated *H19* and *MEG3* DMRs, we observed CpG methylation at all developmental ages in male cKIT$^+$ PGCs. In contrast, maternally methylated DMRs in the testis displayed a sharp reduction in CpG methylation between 16-17 weeks, and for *KCNQ1* methylation was completely lost in the one 20-week sample consented to our study. Analysis of the ovary revealed a significant reduction of CpG methylation by 16.5 weeks at all loci. At paternally methylated DMRs, erasure was complete by 14.5-15 weeks for *H19*, and near complete by 16.5 weeks for *MEG3*.

5hmC is the major methylation species in the common PGC progenitor and is localized to *PEG3* DMR prior to demethylation

Immunofluorescence for 5hmC, the oxidized derivative of 5mC revealed robust nuclear staining in somatic cells at all time points (open arrowheads on Figure 4-3e,f), similar to previous reports in the mouse[18]. However in the germ line 5hmC expression is dynamic, exhibiting punctate nuclear staining in the common PGC progenitor stage (arrows on Figure 4-3e,f), which is lost in OCT4A$^+$ PGCs in the testis from 13.5-16 weeks (Figure 4-3e). Enrichment of 5hmC is again detected in some OCT4A$^+$ PGCs by 17 weeks (Figure 4-3e). In fetal ovaries 5hmC is heterogeneous at 11-19 weeks, being enriched in some but not all OCT4A$^+$ cells (Figure 4-3f).

Bisulfite conversion does not distinguish between 5mC and 5hmC, therefore we used combined glycosylation restriction analysis (CGRA) at the *PEG3* DMR to identify whether 5hmC is enriched at this imprinted locus in PGCs relative to somatic cells or hESCs (Figure 4-3g). We identified 5ghmC at all stages of PGC development in both sexes at the PEG3 DMR. In contrast 5ghmC was not enriched at the PEG3 locus in BJ and H1 cells.

H3K27me3 and H2A.Z are enriched in common PGC progenitor cells

In the mouse, gonadal epigenetic reprogramming of PGCs and imprint erasure occurs from e11.5-e12.5 coincident with global changes in chromatin, including a transient loss of H3K27me3 and a permanent loss of the histone variant H2A.Z[12]. In humans, using immunofluorescence we show that H3K27me3 is enriched in the nucleus of common PGC progenitors in testes from 7-10.5 weeks (Figure 4-4a,c). However, at 11 weeks, the endpoint of the common progenitor stage, H3K27me3 is at or below the level of detection in the majority of $OCT4A^+$ and $VASA^+$ PGCs (Figure 4-4a,c). Interestingly at 17 weeks in testes, H3K27me3 is again observed the nucleus of ~38% $OCT4A^+$ and $VASA^+$ PGCs. In ovaries, H3K27me3 is absent in 50-60% of common PGC progenitors at 6-8.5 weeks (Figure 4-4b,d), after which all PGCs are negative for H3K27me3 (Figure 4-4b,d). Similarly, H2A.Z is enriched in the nucleus of common progenitor stage PGCs at 7-9 weeks in the testis and 7.5 weeks in the ovary (Figure 4-4e,f). However, at the end of the common progenitor stage all PGCs become devoid of H2A.Z until around 17 weeks when H2A.Z reappears in the nucleus of a few $VASA^+$ cells in both sexes (Supplementary Figure 4-S5b,c).

RNA-Seq reveals that $cKIT^+$ PGCs are transcriptionally distinct from hESCs

To generate a comprehensive portrait of cKIT$^+$ PGCs in the fetal testis and ovary, we performed RNA-Sequencing (RNA-Seq) of cKIT$^+$ PGCs sorted at 16-16.5 weeks from fetal testes (n=2), fetal ovaries (n=2) and H1 hESCs sorted with the pluripotent marker TRA-1-60 (n=3). At this developmental time point, male cKIT$^+$ PGCs are initiating imprint erasure, whereas in females some imprinted loci show near complete demethylation (*H19* and *MEG3*). A heatmap of the 5,455 differentially expressed genes in at least one of three pair-wise comparisons is shown in Figure 5a. Pearson Correlation Coefficient analysis showed strong correlations between biological replicates in each group (Figure 4-5b). Gene Ontology (GO) analysis of the 13 differentially expressed gene clusters revealed that male and female cKIT$^+$ PGCs are enriched in GO terms including negative transcription regulation, sex differentiation, and in females, meiosis and germ plasm when compared to hESCs. In contrast, hESCs are enriched in GO terms related to macromolecule biosynthetic processing, RNA processing/splicing and mitosis (Figure 4-5a). Comparing testicular and ovarian cKIT$^+$ PGCs revealed 433 differentially expressed genes, with GO terms such as meiosis, oocyte development and DNA repair. In females this included enrichment in *DAZL*, *VASA*, *ZP3* and *STRA8,* and in males *NANOS2* and *NANOS3* (Figure 4-5c).

Given that 5hmC was detected at 16-16.5 weeks by either CGRA and/or immunofluorescence in cKIT$^+$ PGCs, we also examined expression of *Ten Eleven Translocation (TET)* genes, which are responsible for converting 5mC to 5hmC[19-21] (Figure 4-5c). All three *TET* family members *(TET1-3)* are expressed by male and female PGCs, with a significant enrichment of *TET2*, and reduced expression of *TET1* relative to H1 hESCs. We also evaluated the *DNA Methyltransferases (DNMT) DNMT1, DNMT3A, DNMT3B and DNMT3L.* Male but not female PGCs exhibited a significant decrease in *DNMT1* relative to H1 hESCs. Furthermore, all PGC samples had

reduced expression of *DNMT3A* and *DNMT3B* relative to H1. *AICDA* (also known as *AID*) and *TDG* were expressed at variable levels in H1 and also in PGCs of both sexes.

cKIT/TRA-1-81 positive PGCs generated *in vitro* correspond to immature pre-gonadal PGCs

We next sought to generate cKIT$^+$ PGCs *in vitro* from H1 (XY) and UCLA1 (XX) hESCs[22,23]. In our sorting strategy we incorporated TRA-1-81, as the second marker with cKIT, based on the high OCT4/TRA-1-81 correlation in the human gonad prior to 10 developmental weeks (Supplementary Figure 4-S3d). Given that undifferentiated hESCs expressed detectable levels of germ line genes by RNA-Seq similar to previously reported [24,25], we performed single cell analysis of sorted TRA-1-60$^+$ hESCs (Figure 4-5d,e) and TRA-1-81$^+$/cKIT$^+$ hESCs (Fig 5f,g). As expected, analysis of 100-pooled undifferentiated hESCs resulted in identification of 5/5 or 3/4 PGC signature genes together with *OCT4*. However interrogation at a single cell level revealed that the major PGC determinant *BLIMP1*[26] was rarely expressed, and the majority of PGC signature genes were seldom co-expressed in single cells regardless of sorting strategy.

Next, PGC differentiation was evaluated for up to 16 days by serum-induced differentiation of hESCs; 1) as embryoid bodies (EBs) with and 2) without BMP4 addition, 3) adherent monolayer differentiation on growth factor reduced matrigel (GFR M/G), 4) differentiation on human fetal gonadal stromal cells (hFGSC), and 5) a combination of EB differentiation, followed by plating EBs on hFGSCs (Figure 4-6 and 7). First we verified that TRA-1-81 faithfully reports OCT4 expression upon hESC differentiation for 28 days by flow cytometry using the H1 OCT4-GFP line created by homologous recombination[27] (Figure 4-6a,b). Using the gating strategy shown in Fig 6c (H1 day 9 EBs), we sorted the brightest TRA-1-81$^+$/cKIT$^+$ cells upon *in vitro* differentiation

averaging 0.3% of the live population, with no significant difference in percentage positive cells when comparing differentiation strategy or length of time in differentiation. Our data show that >97% of TRA-1-81^{+}/cKIT^{+} differentiated cells are negative for CD45, excluding the possibility of contamination with cKIT^{+} hematopoietic progenitors[28] (Figure 4-6c). Single cell analysis of cKIT^{+}/TRA-1-81^{+} cells sorted from EBs at day 9 revealed a significant increase in the proportion of cells expressing *BLIMP1* (Fig 6b). We show that the identity of putative PGCs was heterogeneous being either O/B double positive, or O/B/N3 triple positive with no co-expression of *DAZL* or *VASA* (Figure 4-6d and Figure 4-7c). Although rare O/B and O/B/N3 single cells were identified in the undifferentiated state (Fig 5d-f), differentiation resulted in a clear enrichment for both O/B and O/B/N3 cell types. In particular, EB differentiation for 9 days yielded 9-fold enrichment in O/B and 7.5-fold enrichment in O/B/N3 cells relative to cKIT^{+}/TRA-1-81^{+} self-renewing hESCs (quantified in Figure 4-7a,b). Immunofluorescence of day 9 EBs verified that cKIT^{+} cells co-expressed OCT4A, exhibited nuclear localization of BLIMP1 and expressed NANOS3 in the cytoplasm similar to what is observed in cKIT^{+} PGCs from the human fetus (Figure 4-6e,f and Supplementary Figure 4-S3c). Immunofluorescence also revealed that all NANOS3/OCT4A^{+} cells in the day 9 EB were positive for 5mC (Figure 4-6i). Taken together, our data suggests that TRA-1-81^{+}/cKIT^{+}/O/B/N3 cells in EBs correspond to human PGCs prior to gonadal colonization and loss of 5mC.

Using PGC differentiation in EBs for 9 days as a comparison, we show that O/B/N3 putative PGCs are transient being lost by day 15 of EB formation (Figure 4-7b,c). Transferring day 9 EBs to hFGSCs for an additional 7 days (16 days total) was consistent with an increase in survival and/or differentiation of O/B/N3 triple positive cells, however *DAZL* or *VASA* RNA were not induced in the O/B/N3 population (Figure 4-7c). Sustaining the O/B/N3 triple positive population

within the TRA-1-81$^+$/cKIT$^+$ fraction for 15 days was also achieved using adherent differentiation on GFR M/G (Figure 4-7c).

DISCUSSION

By analyzing 134 human embryonic and fetal samples from 6-20 developmental weeks and *in vitro* PGC differentiation from hESCs we propose the following roadmap of human germ line development (Figure 4-8). Our data reveal that the first 16 days of hESC differentiation *in vitro*, either as EBs or as monolayers creates a cKIT$^+$/TRA-1-81$^+$/OCT4A$^+$ PGC population equivalent to a pre-gonadal, PGC with 5mC. After reprogramming 1 (denoted by the global depletion of 5mC from the genome followed by enrichment of H3K27me3[11]), we speculate that DAZL and VASA are next expressed giving rise to the cKIT$^+$/OCT4A$^+$/VASA$^+$ common gonadal PGC progenitors which then embark on reprogramming 2 and uncoupling of cKIT expression from VASA. This uncoupling of germ cell-expressed genes into separate populations was previously reported in second trimester testes and ovaries for OCT4 and VASA protein[29]. Our data is in agreement with cKIT being on the surface of OCT4$^+$/VASA negative cells in the fetal gonad[29]. Furthermore, the SYCP3 staining described here also supports the hypothesis that single VASA$^+$ germ cells in the ovary are first to enter the ovarian reserve in fetal life[29].

Our results show that reprogramming 2 in the cKIT$^+$ lineage follows a protracted series of events that are similar but not identical to the mouse. This begins with the relatively stable wholesale epigenetic loss of H3K27me3 and H2A.Z in the common progenitor followed by either loss of OCT4A or expression in the cytoplasm. Traditionally cytoplasmic localization of OCT4 is due to expression of the OCT4B splice variant[16]. Here we used an antibody that discriminates OCT4A from OCT4B[17] suggesting that OCT4A in PGCs either translocates to the cytoplasm, or is

attenuated there possibly for degradation. The significance of cytoplasmic OCT4A is unknown, but is notably coincident with major global epigenetic changes.

A major event in reprogramming 2 is the erasure of cytosine methylation from DMRs of imprinted genes, which in mice is hypothesized to be active[13]. In the current study we show that 5hmC and the *TET* enzymes are dynamically expressed by cKIT$^+$ human PGCs, as well as additional molecular candidates that could actively modify 5mC/5hmC or remove modified 5hmC from the genome including, *AICDA* and *TDG*[14,19-21,30,31]. In hESCs and fibroblasts where cytosine methylation at imprinted DMRs is stably inherited, 5hmC is not detected at *PEG3* DMR. In contrast in cKIT$^+$ PGCs, where the fate of this locus is demethylation, 5hmC is enriched. Despite this tantalizing correlation, future studies are needed to determine the role of TETs and 5hmC in imprint erasure. One observation from our reference map is that imprint erasure occurs over weeks and in a locus specific manner. This relatively long window for imprint erasure in cKIT$^+$ PGCs stands in contrast to the mouse where erasure at many imprinted loci occur within 24 hours[32]. Therefore, our data support the idea that removal of 5mC from imprinted DMRs in humans may involve diverse, locus-specific and time-dependent strategies.

Using undifferentiated hESCs we show that germ line genes are expressed in a stochastic manner in the undifferentiated state, with only rare undifferentiated hESCs expressing the major PGC determinant *BLIMP1*[26]. However, with *in vitro* differentiation using cKIT with TRA-1-81 we enriched for BLIMP1 expression in cKIT$^+$ cells, called O/B cells that we speculate represent the first lineage restricted PGCs equivalent to e6.25 in mice[33,34]. Differentiation also results in enrichment of O/B/N3 triple positive cKIT$^+$ cells that we speculate represent the next stage in PGC development after O/B and prior to 5mC loss in reprogramming 1. O/B/N3 cells were never

observed in single hESCs sorted for TRA-1-60 and were found in less than 5% of TRA-1-80^{+}/cKIT^{+} sorted undifferentiated hESCs. The fact that at a single cell level *VASA* and *DAZL* were never co-expressed with O/B/N3 in TRA-1-81^{+}/cKIT^{+} cells with differentiation does not refute previous findings using *DAZL* and *VASA* as markers to define germ line identity[4,24,35-41]. On the contrary, we propose that sorting for TRA-1-81^{+}/cKIT^{+} specifically enriches for newly specified germ line cells prior to *DAZL* and *VASA* expression and that acquisition of this immature cell can be achieved regardless of differentiation strategy.

In conclusion, we show that accurate interpretation of *in vitro* differentiation requires not only a detailed understanding of the human counterpart, but also analysis at a single cell level to confirm molecular identity and rule out stochastic gene expression. In the long term for the field to move forward, functional assays to determine human germ line quality are urgently required. One possibility is using nonhuman primate hPSCs where transplantation of *in vitro* derived germ cells is ethically possible. Alternatively, methods for culturing endogenous PGCs in a format that promotes self-renewal and/or differentiation to gonocytes, gonia and meiotic cells rather than EGCs is also needed. The human germ cell lineage is particularly challenging to study due to lack of functional assays to test germ cell identity and quality, therefore the generation of robust molecular maps as described here are the first steps to unveiling this important lineage.

MATERIALS AND METHODS

Human Fetal Samples

Fetal testes and ovaries were acquired following elected termination and pathological evaluation for this research program only after UCLA-IRB review which deemed the project exempt under

45 CRF 46.102(f). The majority of samples (112) were obtained from the University of Washington Birth Defects Research Laboratory (BDRL), under the regulatory oversight of the University of Washington IRB approved Human Subjects protocol combined with a Certificate of Confidentiality from the Federal Government. BDRL harvests the embryonic and fetal tissues (testes and ovaries) and ships them overnight for immediate processing in our laboratory in Los Angeles. In rare cases (21) we obtained de-identified fetal gonads from the UCLA Translational Pathology Core Laboratory and the UCLA Gene and Cellular Core Laboratory (1). All consented material was anonymous and carried no personal identifiers. The BDRL estimates developmental age by prenatal intakes, foot length, Streeter's stage and crown-rump length. Samples with a documented birth defect or chromosomal abnormality were excluded from our study. At UCLA Translational Pathology Core Laboratory and the UCLA Gene and Cellular Core Laboratory, developmental age is calculated by recall of last menstrual period minus 2 weeks. After completing this study we excluded n=2 fetal testes acquired from UCLA Translational Pathology Core Laboratory that appeared to be older than the developmental age provided after comparing the outcome to the more accurate staging used for the 112 samples acquired from the University of Washington BDRL.

Combined Glycosylation Restriction Analysis (CGRA) at PEG3 DMR

For CGRA analysis at the PEG3 DMR, DNA was extracted using the Quick-gDNA MiniPrep Kit (Zymo Research) according to manufacturers' instructions. CGRA was next performed using the EpiMark 5-hmC and 5-mC Analysis Kit (NEB) according to manufacturers' instructions followed by q-RT- PCR for PEG3 DMR DNA using primers PEG3 Forward: 5'-CCACCTGCAGCCACTTC-3' and PEG3 Reverse: 5'- AGTTGGTTGGGCGAGACAAG-3' with SYBR Green Master Mix (Applied Biosystems) at Tm=65 ^{o}C on a Bio-rad MyiQ Thermal Cycler (Bio-rad).

Cell culture

Human ESC lines H1 (WA01, 0043, 46XY) and UCLA1 (0058, 46XX) were maintained under self-renewal conditions on mouse embryonic fibroblast (MEF) layer in DMEM:F12 (Gibco BRL), 20% KnockOut Serum (Gibco BRL), 1% nonessential amino acids (NEAA, Gibco BRL), 1 mM L-glutamine (Gibco BRL), 0.1 mM β-mercaptoethanol (Gibco BRL), and 10ng/ml of basic fibroblast growth factor (FGF) from the Biological Resources Branch of the Frederick National Laboratory for Cancer Research. Undifferentiated hESC colonies were passaged every 7 days and maintained as previously described[4]. For EB differentiation, on day four after passage undifferentiated H1 and UCLA1 cells were treated with Collagenase Type IV (Gibco BRL) (1mg/ml) for 45 min at 37 ^{o}C. The colonies that dissociated were lifted by gentle pipetting and plated overnight in ultra low attachment 6-well plates (Corning Incorporated, Corning NY) in mTeSR1 medium (StemCell Technologies) supplemented with 10nM ROCK Inhibitor (HA-1077, Sigma-Aldrich). 24 hrs after initial plating medium was replaced by differentiation medium, DMEM:F12 supplemented with 20% FBS, 0.1 mM nonessential amino acids, 0.1 mM β-mercaptoethanol, 1 mM L-glutamine (all reagents from Gibco BRL). Medium was changed every third day. Addition of 50 ng/mL carrier free BMP4 (R&D Systems) was used where necessary. Adherent differentiation was performed on plates coated with growth factor reduced matrigel (BD Pharmigen) in differentiation medium, changed every two days. For differentiation on FGSCs, differentiation medium was changed every two days. FGSCs were cultured as previously described[4]. For all experiments, hESCs were used between passages 8 and 45. All hESC experiments were conducted with prior approval from the UCLA Embryonic Stem Cell Research Oversight Committee. BJ fibroblast somatic cells were cultured in minimum essential medium (MEM) with Earle's salt (Gibco BRL) and 1 mM L-glutamine, 10% FBS (Gibco BRL), 1% NEAA and 1 mM sodium pyruvate (Gibco BRL). HEK 293 cells were maintained on gelatin-coated plates in DMEM High Glucose (Gibco BRL) supplemented with 1 mM L-glutamine, 10%

FBS (Gibco BRL), 1% NEAA and 1 mM sodium pyruvate (Gibco BRL). BJ and HEK 293 cells were passaged using 0.25% trypsin (Gibco BRL) every 5 days.

FACS and Flow cytometry

Cells from EBs were dissociated in 0.25% Trypsin-EDTA (Gibco BRL) supplemented with 2% Chicken serum (Gibco BRL) at 37° C for 30 min and collected by centrifugation at 1.000 rpm for 5 min. For adherent differentiation, human fetal gonads and HEK 293, cells were dissociated with 0.25% Trypsin-EDTA for 5 min at 37 ^{o}C. Dissociated cells were incubated in 1% BSA in PBS containing primary antibodies on ice for 20 min. Primary antibodies used for flow analyses were, cKIT-APC conjugated (1:100, 550412; BD Biosciences), TRA-1-81 (1:100, 14-8883; eBioscience), TRA-1-60 (1:100, 14-8863; eBioscience), and CD45-PE conjugated (1:50, PN IM1833; Beckman-Coulter). Cells were then washed and incubated with FITC-conjugated (Jackson ImmunoResearch) or Pe-Cy7 conjugated secondary antibody (BD Biosciences) on ice for another 20 min. Cells were washed again and incubated with 1% BSA in PBS for 5 min on ice. To ensure single cell separation before flow analysis or FACS sorting cells were passed through a 40 uM filter (BD Biosciences). As a viability dye, for in vitro differentiation 7-AAD (1:50, BD Pharmigen) was used, for gonadal samples DAPI at 1:1000 dilution (10 ug/ml, Sigma) was used. Analysis was performed using LSR II (Becton Dickinson) and FlowJo software (Tree Star Inc).

Immunofluorescene

Immunofluorescence of fetal gonads from 6 to 19 weeks of gestation was performed as previously described[4]. Dilutions and catalogue numbers of primary antibodies used were: cKIT (1:100, A4502; DAKO), OCT4A (1:100, (N-19)-sc-8628; Santa Cruz), VASA (1:400, AF2030; R&D Systems), H3K27me3 (1:800, 05-1339; Millipore), H2A.Z (1;100, ab4174 Abcam); 5mC (1:100, AMM99021; Aviva), NANOS3 (1:300, ab70001; Abcam), SYCP3 (1:50, NB300-232;

Novus Biologicals), BLIMP1 (1:100, 9115; Cell Signaling), TRA-1-81 (1:100; 14-8883; eBioscience), SSEA1 (1:50, MC-480; Jackson ImmunoResearch) and 5hmC (1:100, 39769; Active Motif). For 5hmC, a denaturing step was added to the staining procedure as previously described[43]. All samples were incubated with primary antibodies overnight at 4 ^{o}C. Sections were washed, incubated with FITC/TRITC conjugated secondary antibodies (Jackson ImmunoResearch) for 30 min and mounted in Prolong Antifade Reagent with DAPI (Invitrogen). Samples were imaged on a Zeiss Axio Imager (Zeiss) using Axio Vision 4.8 Software (Zeiss).

Single-cell Real Time RT-PCR

Single cells were sorted with a BD ARIA cell sorter equipped for Biosafety Level 2 (BSL2) sorting, and single cell analysis was performed as described[44].

RNA extraction and RNA-Seq Library generation

Cells were sorted directly in 75ul RLT buffer (Qiagen) and RNA was extracted using the RNeasy Micro Kit (Qiagen) according to manufacturer's instructions. RNA was amplified and converted to ssDNA using the WT-Ovation RNA Amplification System (Nugen). dsDNA was generated from ssDNA using the WT-Ovation Exon Module (Nugen) according to the manufacturers' instructions. dsDNA was sonicated to DNA fragments within a 200-500bp range. Subsequently the libraries were generated starting from 30ng DNA using the TruSeq DNA Sample Preparation Kit (Illumina) according to manufacturers instructions. Libraries were run using 50bp single-end reads on the HiSeq 2000 System (Illumina).

RNA-Seq analysis

We performed single-end RNA-Seq analysis and obtained 116,399,186 (UW30) and 153,982,474 (UW99) total reads of which 35,497,900 and 37,907,565 respectively were uniquely mapped to the reference genome (UCSC assembly hg19) for the cKIT^{+} fetal testis

libraries. For the cKIT$^+$ fetal ovary libraries, we obtained 211,659,065 (UW92) and 102,837,931 (UW98) total reads, of which 93,766,963 and 44,784,216 reads respectively were uniquely mapped. For the three H1 hESC replicates we obtained 161,037,803 (A), 133,355,463 (B) and 104,997,235 (C) total reads, of which 39,339,764, 38,571,046 and 37,164,448 reads respectively were uniquely mapped. Reads were mapped to the reference genome using the gapped aligner Tophat (version 1.3.0) allowing up to 2 mismatches. The human gene model annotation (version of Homo_sapiens.GRCh37.59) was downloaded from the Ensembl database and supplied to Tophat. Gene and transcript expression levels were quantified using Cufflinks (version 1.0.3) in the FPKM unit (Fragments Per Kilobase of exon per Million fragments mapped) together with confidence intervals. Cufflinks ran in the default parameters except that the annotated gene set was supplied using the -G option. Raw read count for each gene and transcript was measured using customized scripts written in Perl. RNA Sequencing data are available in GEO with accession number GSE39821.

Differential expression testing

Differential expression analysis was performed using the R package, DESeq and edgeR. Raw read counts were employed and modeled as negative binomial distributed because our data set contains biological replicates. We filtered out lowly expressed genes and transcripts by only keeping those that have at least one count per million in samples. The multiple testing errors were corrected by the false discovery rate (FDR). Besides the cutoff as adjusted at $P < 0.05$, we also adopted an additional cutoff set as the expression ratio of above two-fold changes in expression values. In summary, we considered genes as differentially expressed if: 1) the adjusted p-value was less than 0.05; 2) the expression ratio between two conditions was above two-fold; 3) agreement between DESeq and edgeR. The heat map in Fig. 5c was generated using the default settings with the function "heatmap.2" in the "gplots" package of R. RPKM

values are centered and scaled by subtracting the mean of the row from every RPKM value and then dividing the resulting RPKM values by the standard deviation of the row.

Bisulfite Sequencing

Genomic DNA was isolated from sorted samples using the Quick-gDNA MiniPrep Kit according to manufacturer's insturctions (Zymo Research). Bisulfite conversion was performed with the EZ DNA methylation kit (Zymo Research) as previously described[44]. Primers for PEG3[45], for KCNQ1 KvDMR1[46], for MEG3 IG-DMR[47] and for H19[48] were described elsewhere.

Statistical analyses and graphs

All data are expressed as mean ± SEM and statistical analyses were performed using Mann-Whitney test. Graphs were generated using GraphPad Prism 5 (GraphPad Software, Inc., La Jolla, CA).

ACKNOWLEDGEMENTS

The authors would like to thank Angela Chen, MD and the UCLA Translational Pathology Core Laboratory and the UCLA Gene and Cellular Core Laboratory for some of the gonadal samples used in this study. We also thank Joseph Hargan-Calvopina, Marisabel Oliveros-Etter and Silvia Diaz-Perez for critical reading of the manuscript, Felicia Codrea and Jessica Scholes for FACS and Steven Peckman from the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research for critical assistance with human subject and embryonic stem cell review. This work was supported primarily by fund number 1R01HD058047 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) (ATC), as well as the

Iris Cantor-UCLA Women's Health Pilot Project (ATC) and 1P01GM081621 from NIGMS. The Laboratory of Developmental Biology, University of Washington, Seattle is supported by NIH Award Number 5R24HD000836 from the NICHD. Human fetal tissue requests can be made to: bdrl@u.washington.edu.

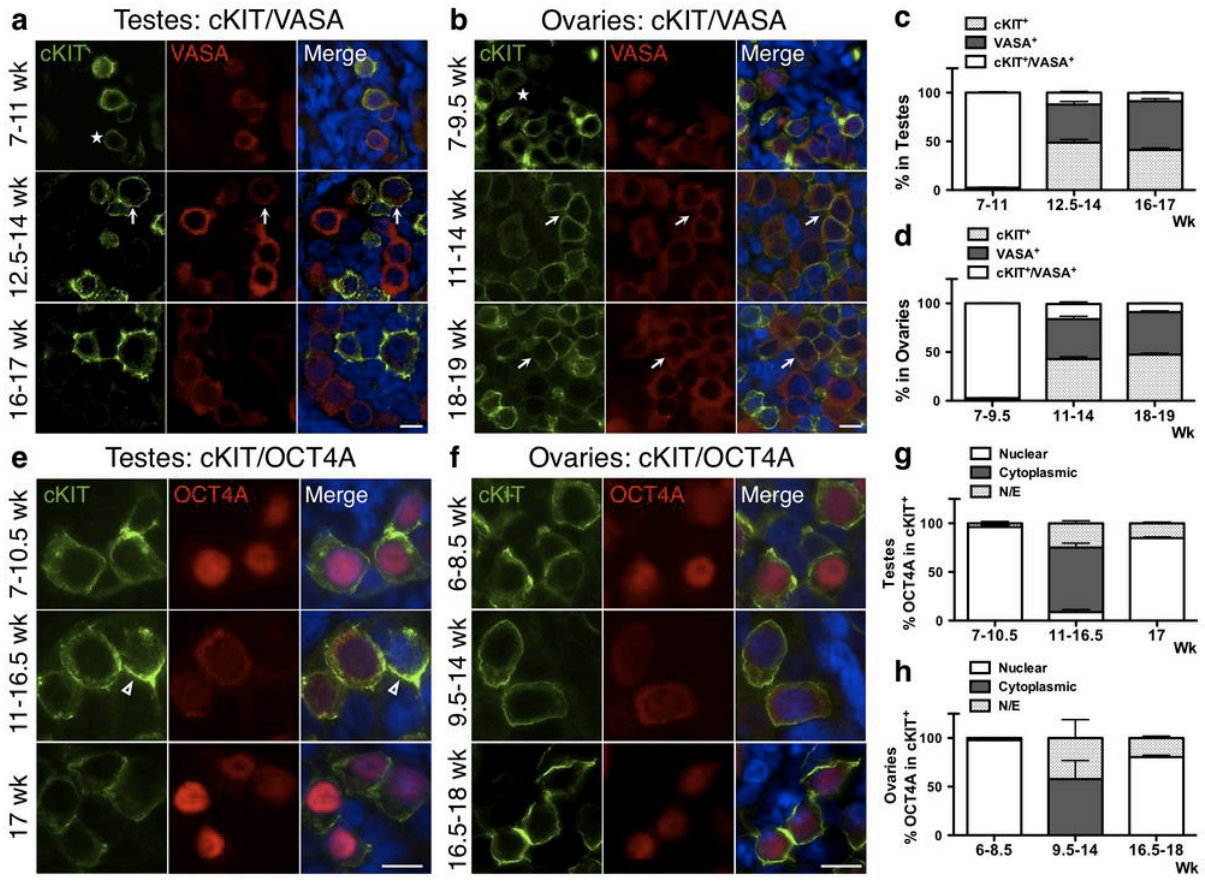

Figure 4-1. The dynamics of cKIT, OCT4A and VASA expression in the fetal gonad.

(a,b) Representative immunofluorescence images of cKIT with VASA at the developmental weeks indicated. Asterisks indicate cKIT dim cells. (a) Shown is a 10-week testis for 7-11wk (n=5), a 13.5-week testis for 12.5-14wk (n=3) and a 16-week testis for 16-17wk (n=2). (b) Shown is a 7-week ovary for 7-9.5wk (n=4), an 11-week ovary for 11-14wk (n=3), and an 18-week ovary for 18-19wk (n=2). (c,d) Quantification of $cKIT^+$, $VASA^+$ and $cKIT^+/VASA^+$ cells (arrows in a,b). (c) In testes, 9 optic fields were counted at 7-11wk (n=5), 9 optic fields at 12.5-14wk (n=3) and 7 optic fields at 16-17wk (n=2). (d) In ovaries, 7 optic fields were counted at 7-9.5wk (n=4), 8 optic fields at 11-14wk (n=3) and 7 optic fields at 18-19wk (n=3). (e,f) Representative immunofluorescence images of cKIT with OCT4A at the developmental weeks indicated. (e) Shown is a 10-week testis for 7-10.5wk (n=5), a 13.5-week testis for 11-16.5wk (n=4) and a 17-week testis (n=1). (f) Shown is an 8-week for 6-8.5wk (n=3), an 11-week for 9.5-14wk (n=3) and an 18-week for 16.5-18wk (n=2). (g,h) Quantification of nuclear or cytoplasmic localization of OCT4A in $cKIT^+$ cells. (g) In testes, 6 optic fields were counted at 7-10.5wk (n=5), 8 optic fields at 11-16.5wk (n=4) and 6 optic fields at 17wk (n=1). (h) In ovaries, 6 optic fields were counted at 7-8.5wk (n=3), 8 optic fields at 9.5-14wk (n=3) and 9 optic fields at 16.5-18wk (n=2). For immunofluorescence, nuclei were counterstained with DAPI (blue) and scale bars represent 10 um. All data are expressed as mean ± SEM. Abbreviations: wk= week, N/E= Not Expressed.

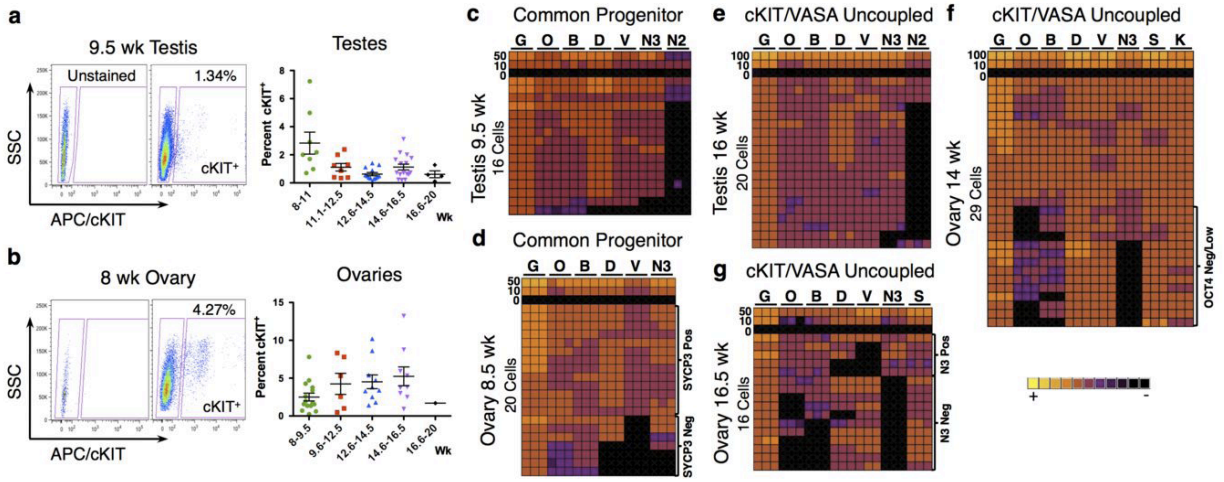

Figure 4-2. Molecular characterization of cKIT$^+$ PGCs from 7-20 developmental weeks.

(a,b) Gating strategy for sorting cKIT$^+$ cells with an APC conjugated anti-human cKIT primary antibody against side scatter (SSC). (a) Shown is a 9.5-week testis and (b) an 8-week ovary. Also shown is the percent of cKIT$^+$ cells sorted from the live fraction of testes in (a) and ovaries in (b) at 8-20 developmental weeks (wk). Each data point represents a single sample (biological replicate). All data are represented as mean ± SEM. (c-g) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3), *cKIT* (K) *NANOS2* (N2) and *SYCP3* (S) in triplicate (columns) in 100, 50, 10, 0 or single sorted cKIT$^+$ cells (rows).

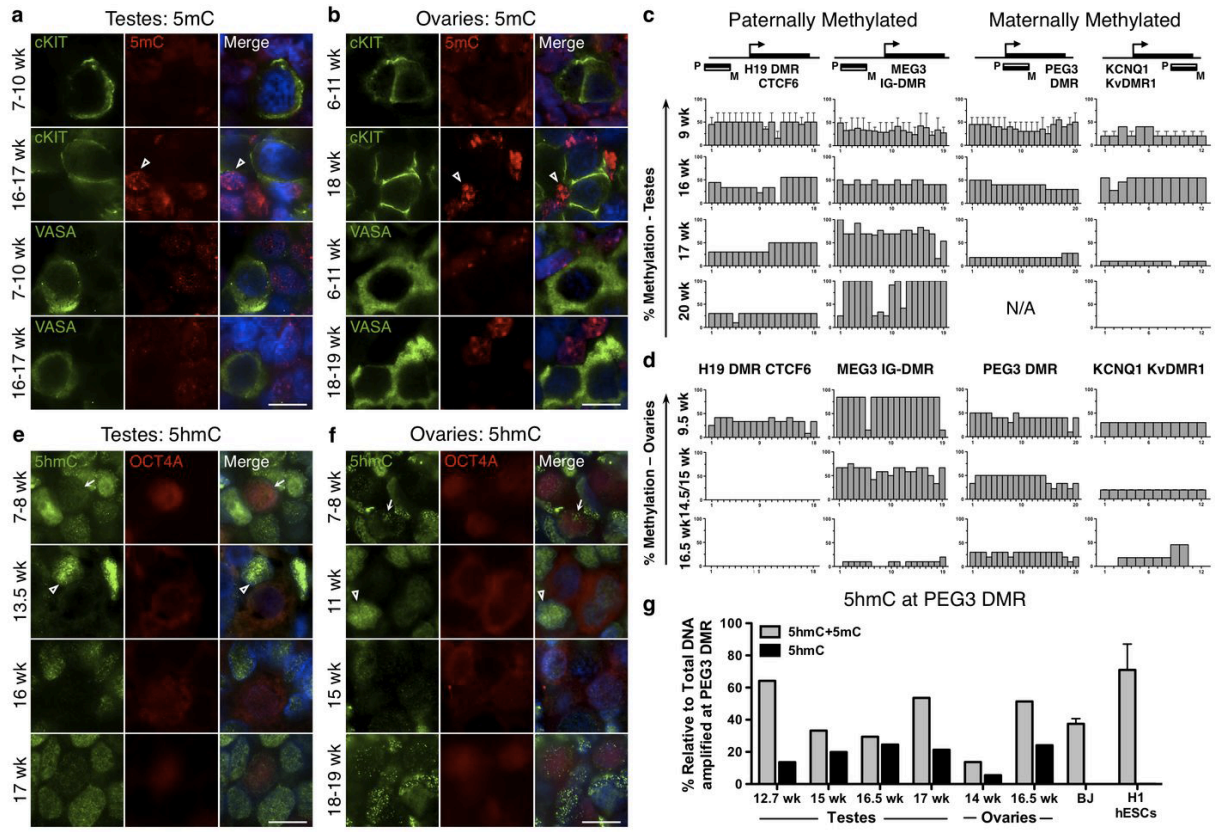

Figure 4-3. Global loss of 5mC precedes loss of 5hmC.

(a,b) Representative immunofluorescence images shown of 5mC with cKIT or with VASA in (a) testes and (b) ovaries at the developmental weeks indicated. Open arrowheads indicate 5mC signal in somatic cells. (a) Shown is a 10-week testis for 7-10wk (n=3) and a 17-week testis for 16-17wk (n=3). (b) Shown is an 8-week ovary for 6-11wks (n=3) and an 18-week ovary for 18-19wk (n=2). (c,d) BS-PCR analysis of H19, MEG3, PEG3 and KCNQ1 in cKIT^{+} PGCs sorted from (c) testes at 9wk (n=2) and at 16, 17 and 20 weeks, and (d) ovaries at 9.5, 14.5, 15 and 16.5 weeks. (e,f) Representative immunofluorescence images of 5hmC with OCT4A in (e) testes and (f) ovaries at the developmental stages indicated in weeks. Arrows indicate 5hmC signal in PGCs, open arrowheads indicate 5hmC signal in somatic cells. (e) Shown is an 8-week testis for 7-8wk (n=2), a 13.5-week testis (n=1) a 16-week testis (n=2) and a 17-week testis (n=1). (f) Shown is an 8-week ovary for 7-8wk (n=2), an 11-week ovary, (n=1), 15-week ovary (n=1) and an 18-week ovary for 18-19 (n=2). (g) CGRA of the PEG3 DMR showing the percent of total methylation (5mC+5hmC) or 5hmC alone, relative to total amplified DNA (uncut) at the PEG3 DMR. DNA from BJ fibroblast and H1 hESCs were used as a negative control (for each, n=2 biological replicates). For immunofluorescence analysis nuclei were counterstained with DAPI (blue). Scale bars represent 10 um. All data are represented as mean ± SEM. Abbreviations: N/A= Not Amplified, wk= week.

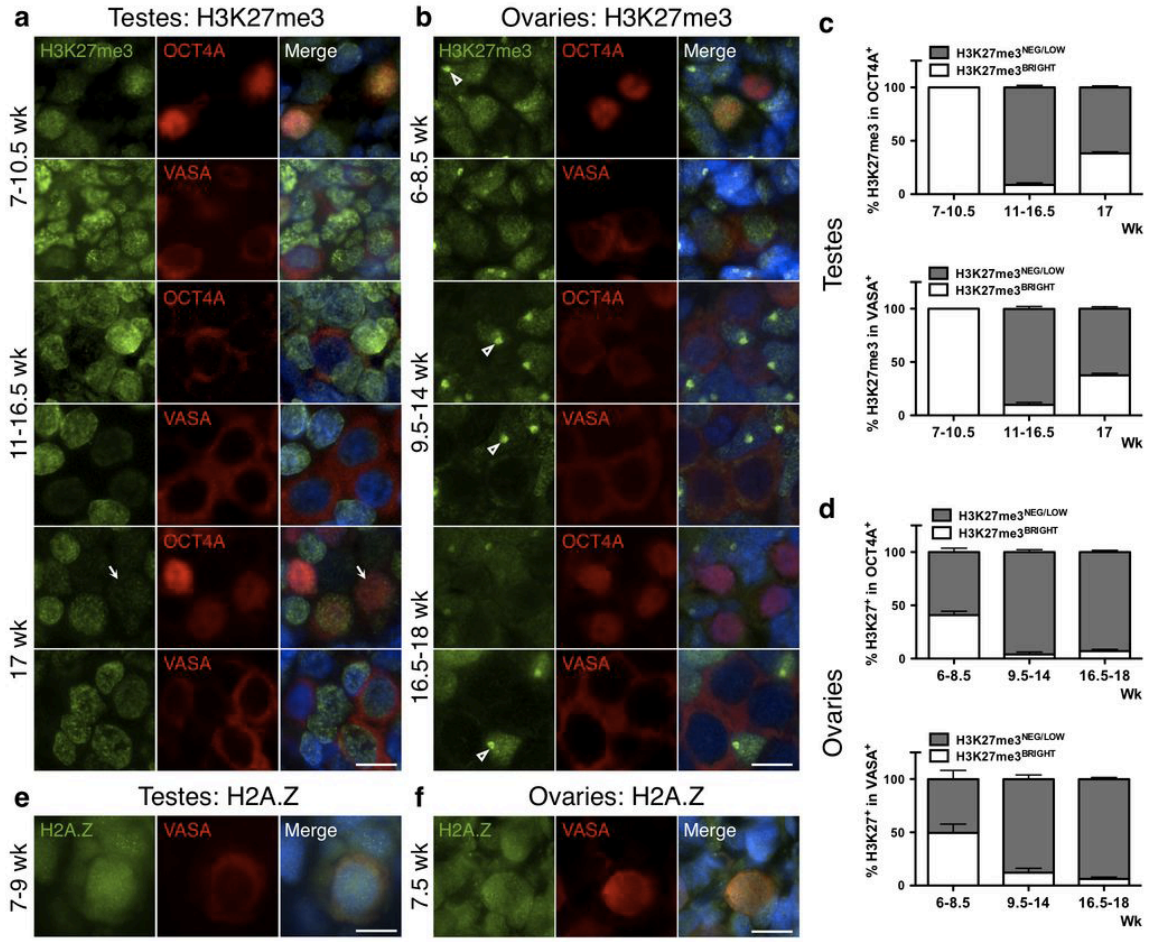

Figure 4-4. Epigenetic reprogramming of H3K27me3 and H2A.Z occurs in the common PGC progenitor.

(a,b) Representative immunofluorescence images of H3K27me3 with OCT4A or VASA in (a) testes from 7-17 weeks, and (b) ovaries from 6-18 weeks. (a) Shown is a 10.5-week testis for 7-10.5wk (n=3), a 16.5-week testis for 11-16wk (n=7) and a 17-week testis (n=1). Arrow indicates PGC nucleus with H3K27me3^{Low} levels at 17wk relative to the intensity of staining in the somatic neighbors in the same section (b) Shown is a 7-week for 6-8.5wk (n=3), an 11-week for 9.5-14wk (n=4), and an 18-week for 16.5-18wk (n=3). Open arrowhead indicates strong H3K27me3 accumulation that is indicative of X chromosome inactivation[42]. (c,d) Quantification of H3K27me3 in OCT4A^{+} or VASA^{+} germ cells in (c) testes and (d) ovaries, at the developmental ages indicated. (c) In testes for quantification in OCT4A^{+}, 6 optic fields were counted at 7-10.5wk (n=3), 14 optic fields at 11-16.5wk (n=7) and 6 optic fields at 17wk (n=1). For quantification in VASA^{+}, 6 optic fields were counted at 7-10.5wk (n=3), 10 optic fields at 11-16.5wk (n=7) and 6 optic fields at 17wk (n=1). (d) In ovaries, for quantification in OCT4A^{+}, 6 optic fields were counted at 6-8.5wk (n=3), 10 optic fields at 9.5-14wk (n=4) and 8 optic fields at 16.5-18wk (n=3). For quantification in VASA^{+}, 4 optic fields were counted at 6-8.5wk (n=3), 10 optic fields at 9.5-14wk (n=4) and 7 optic fields at 16.5-18wk (n=3). (e,f) Representative immunofluorescence images of H2A.Z with VASA in testes (e) from 7-9 weeks, and ovary (f) at 7.5 weeks. (e) Shown is a 9-week testis for 7-9wk (n=2) and (f) at 7.5-week ovary (n=2). For immunofluorescence, nuclei were counterstained with DAPI (blue). Scale bars represent 10 um. Data are represented as mean ± SEM. Abreviations: wk= weeks.

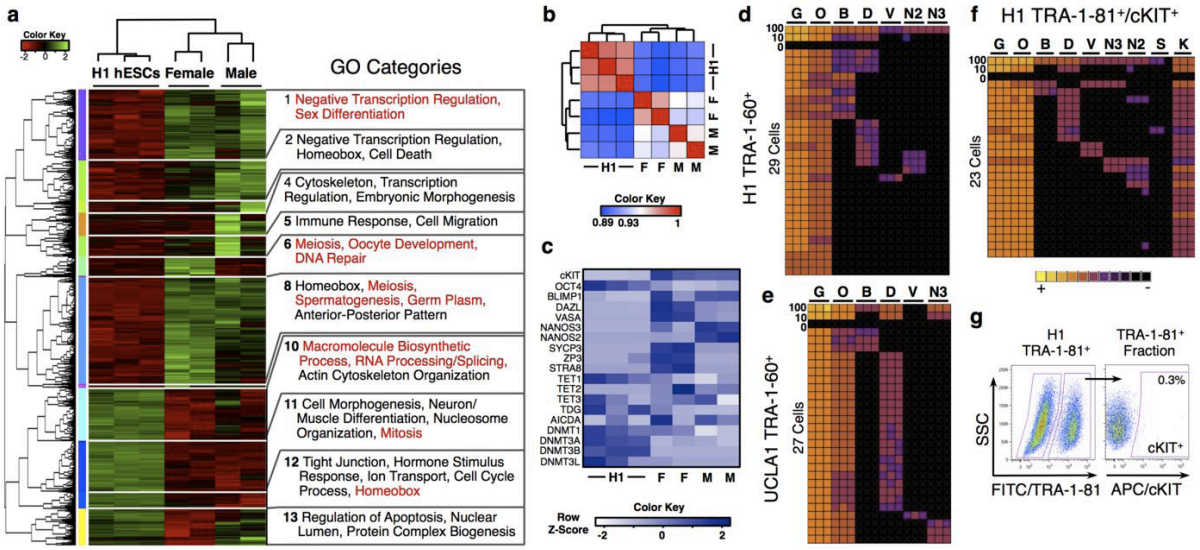

Figure 4-5. RNA-Seq reveals the transcriptional identity of cKIT$^+$ PGCs. Single cell analysis of hESCs shows stochastic expression of germ line genes.

(a) Heat map of 5,455 differentially expressed genes (p value < 0.05) in at least one of three comparisons (male cKIT$^+$ vs H1 hESCs; female cKIT$^+$ vs H1 hESCs; male cKIT$^+$ vs. female cKIT$^+$). The enriched GO terms in the 13 resulting clusters are shown. (b) Heat map of Pearson Correlation Coefficient scores between hESCs and cKIT$^+$ male and female PGCs. (c) Heat map of FKPM values for selected genes in hESCs and cKIT$^+$ male and female PGCs. Abbreviations: M= Male, F= Female. (d,e) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3) and *NANOS2* (N2) for H1 hESCs in triplicate (columns) in 100, 10, 0 or single TRA-1-60$^+$ cells (rows) sorted from H1 (d) and UCLA1 (e) hESCs. (f) Heat map as in d and e plus *cKIT* (K) for TRA-1-81$^+$/cKIT$^+$ H1 hESCs. (g) Gating strategy to sort TRA-1-81$^+$/cKIT$^+$ cells from the H1 hESC line. cKIT$^+$ cells are gated from the TRA-1-81$^+$ fraction, using a FITC secondary antibody against side scatter (SSC).

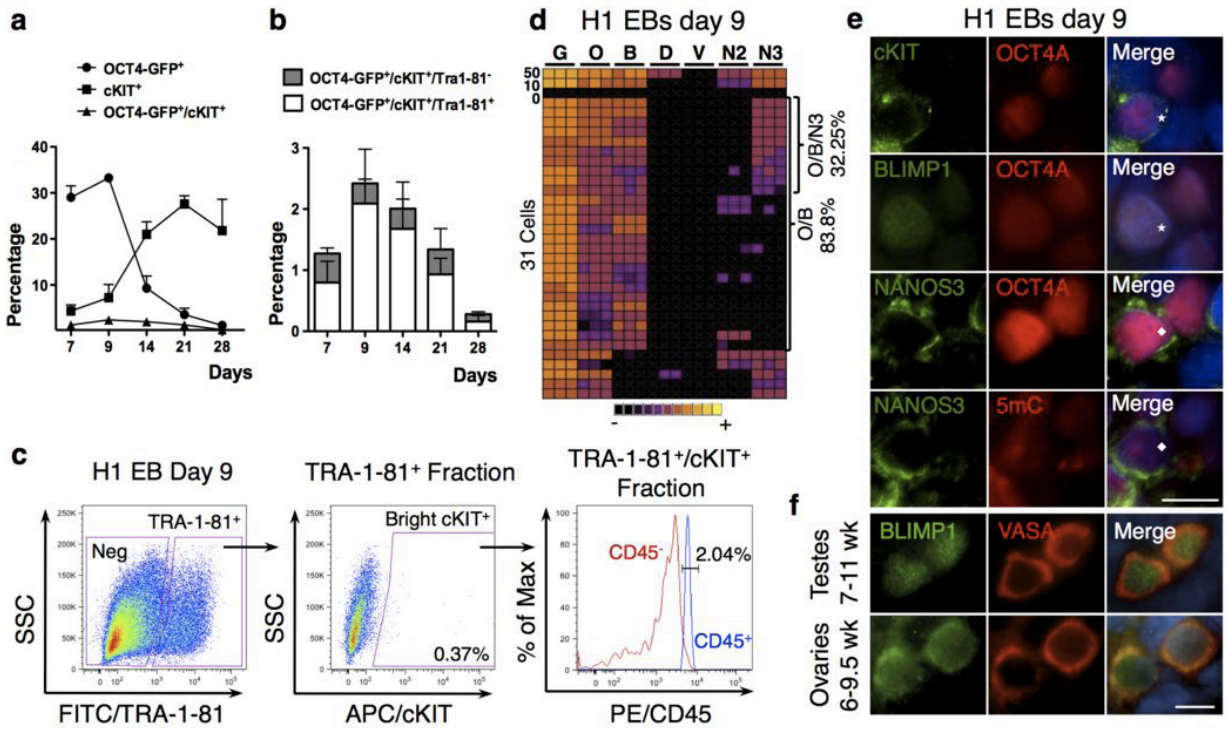

Figure 4-6. ***In vitro*** **hESC differentiation generates rare germ line progenitors that are cKIT/TRA-1-81 positive.**

(a) Percentage of OCT4-GFP$^+$, cKIT$^+$ and total OCT4-GFP$^+$/cKIT$^+$ generated upon adherent differentiation for the indicated time points. (b) Percentage of TRA-1-81$^+$ and TRA-1-81$^-$ cells within the OCT4-GFP$^+$/cKIT$^+$ population generated by adherent differentiation of H1 OCT4-GFP for the indicated time points. TRA-1-81 is co-expressed by the majority of cKIT/OCT4-GFP double positive cells upon hESC differentiation. (c) Gating strategy to sort cKIT$^+$/TRA-1-81$^+$ cells (shown are H1-EBs differentiated for 9 days). cKIT$^+$ cells are gated from the TRA-1-81$^+$ fraction. Flow cytometry for CD45 in the TRA-1-81$^+$/cKIT$^+$ cells reveals <3% contamination by CD45 positive cells. (d) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS2* (N2) and *NANOS3* (N3), in triplicate (columns) in 50, 10, 0 or single cells (rows) for TRA-1-81$^+$/cKIT$^+$ sorted from H1 day 9 EBs. (e) Immunofluorescence of cKIT, BLIMP1 and NANOS3 with OCT4A, and of NANOS3 with 5mC on EBs differentiated for 9 days from H1 hESCs. Staining is performed on adjacent sections and asterisk and diamond indicate the same cell. (f) Representative immunofluorescence images of BLIMP1 with VASA in testes from 7-10.5 weeks (n=3), shown is a 10.5-week testis and ovaries from 6-8 weeks (n=3), shown is an 8-week ovary. Nuclei were counterstained with DAPI (blue). Scale bars represent 10 um. Abbreviations: wk= week. All data are mean ± SEM.

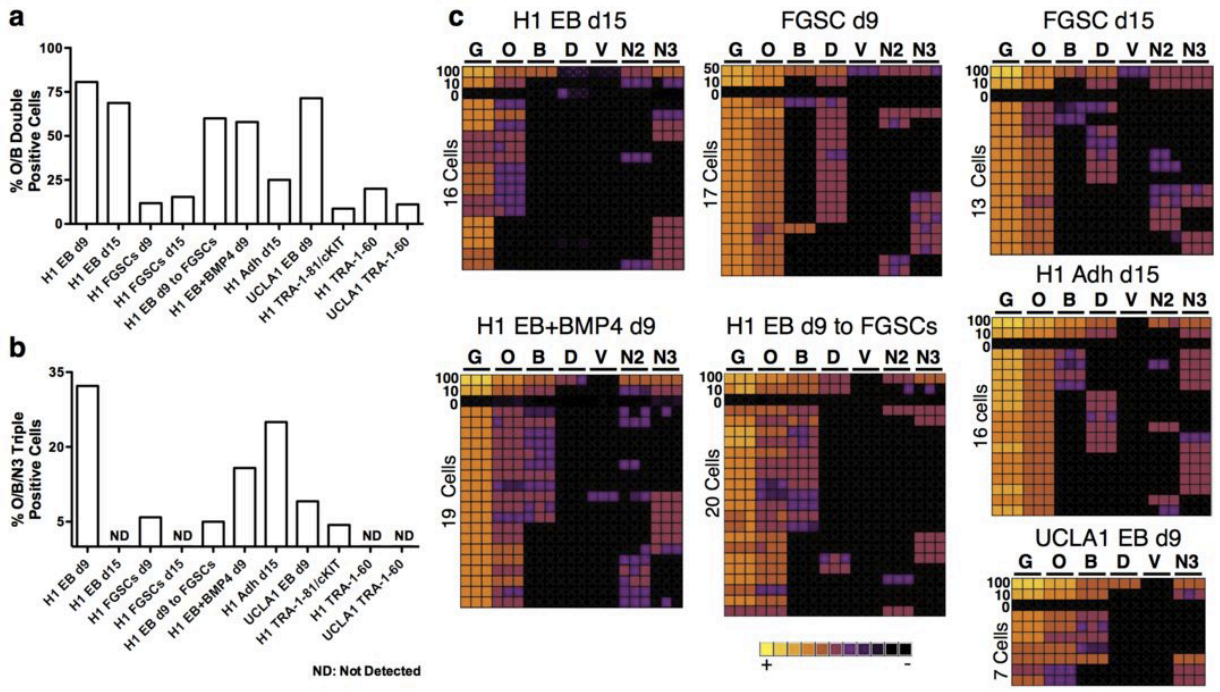

Figure 4-7. ***In vitro*** **PGC differentiation from hESCs using five alternate differentiation techniques.**

(a,b) Comparison of (a) the percent of O/B double positive single cells, and (b) O/B/N3 triple positive within the TRA-1-81^{+}/cKIT^{+} sorted fraction from five alternate differentiation techniques. (c) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3) and *NANOS2* (N2) for H1 in triplicate (columns) in 100, 50, 10, 0 or single TRA-1-81^{+}/cKIT^{+} cells (rows) sorted from H1 and UCLA1 hESCs using five alternate differentiation strategies.

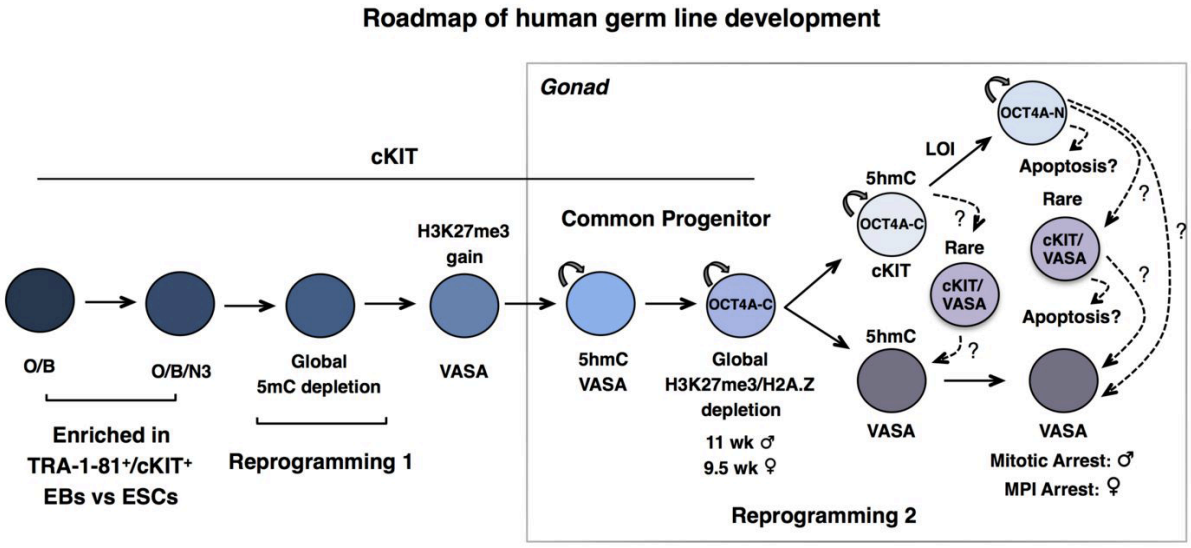

Figure 4-8. Summarized roadmap of human germ line development.

Reprogramming 1, occurs prior to 6-7 developmental weeks and is characterized by global loss of 5mC from PGC DNA. Reprogramming 2 begins in the common PGC progenitor stage after acquisition of H3K27me3 (10.5 weeks in testes and 8.5 weeks in ovaries) and involves global loss of H3K27me3 and H2A.Z followed by imprint erasure in cKIT$^+$ PGCs more than 1 month later. *In vitro* hESC differentiation using TRA-1-81$^+$/cKIT$^+$ sorting generates a rare cKIT$^+$ PGC population that is O/B double positive or O/B/N3 triple positive and corresponds to newly specified PGCs, prior to reprogramming 1. Abbreviations: O/B = *OCT4*/*BLIMP1*, O/B/N3 = *OCT4*/*BLIMP1*/*NANOS3*, LOI= Loss Of Imprinting, MPI= Meiotic Prophase I.

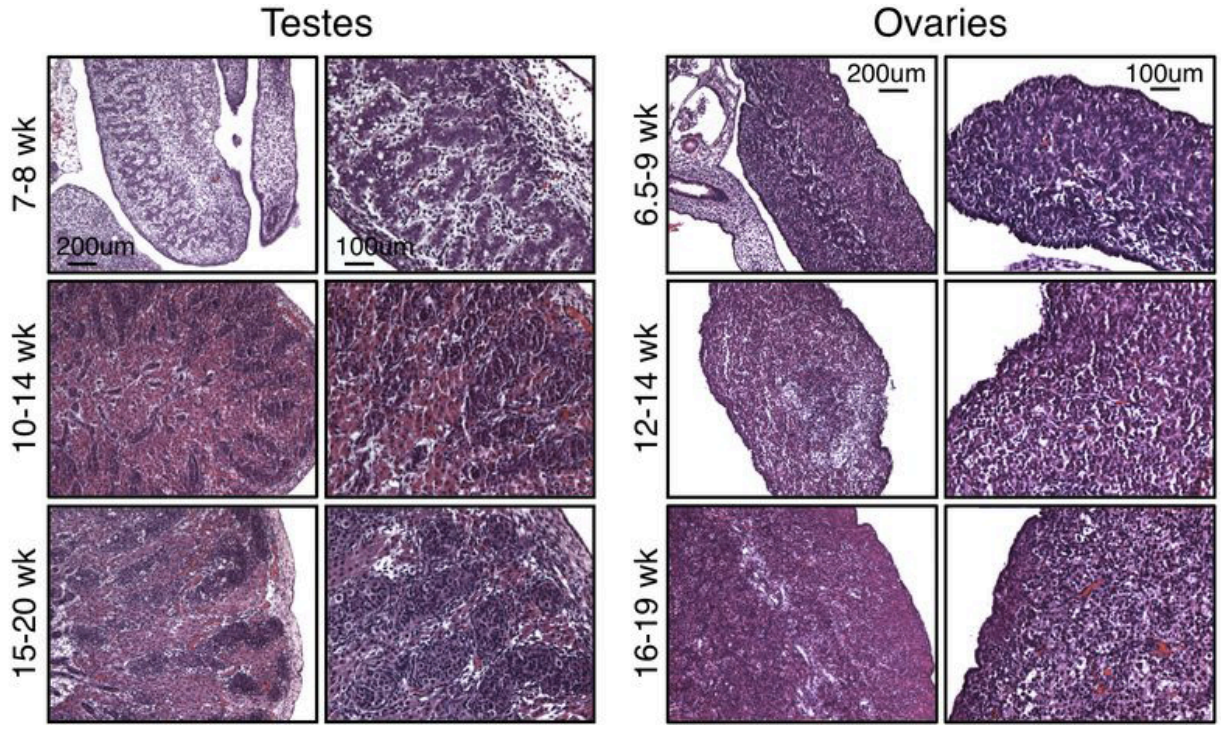

Figure 4-S1. Morphological characteristics of fetal testes and ovaries from 6.5-20 developmental weeks.

Representative images of Hematoxylin and Eosin staining of testes and ovaries over the developmental time period analyzed in this project. In testes, shown is an 8-week for 7-8wk (n=2), a 10-week for 10-14wk (n=3) and a 15-week for 15-20wk (n=5). In ovaries, shown at 6.5-week for 6.5-9wk (n=4), a 14-week for 12-14wk (n=2) and an 18-week for 16-19wk (n=3).

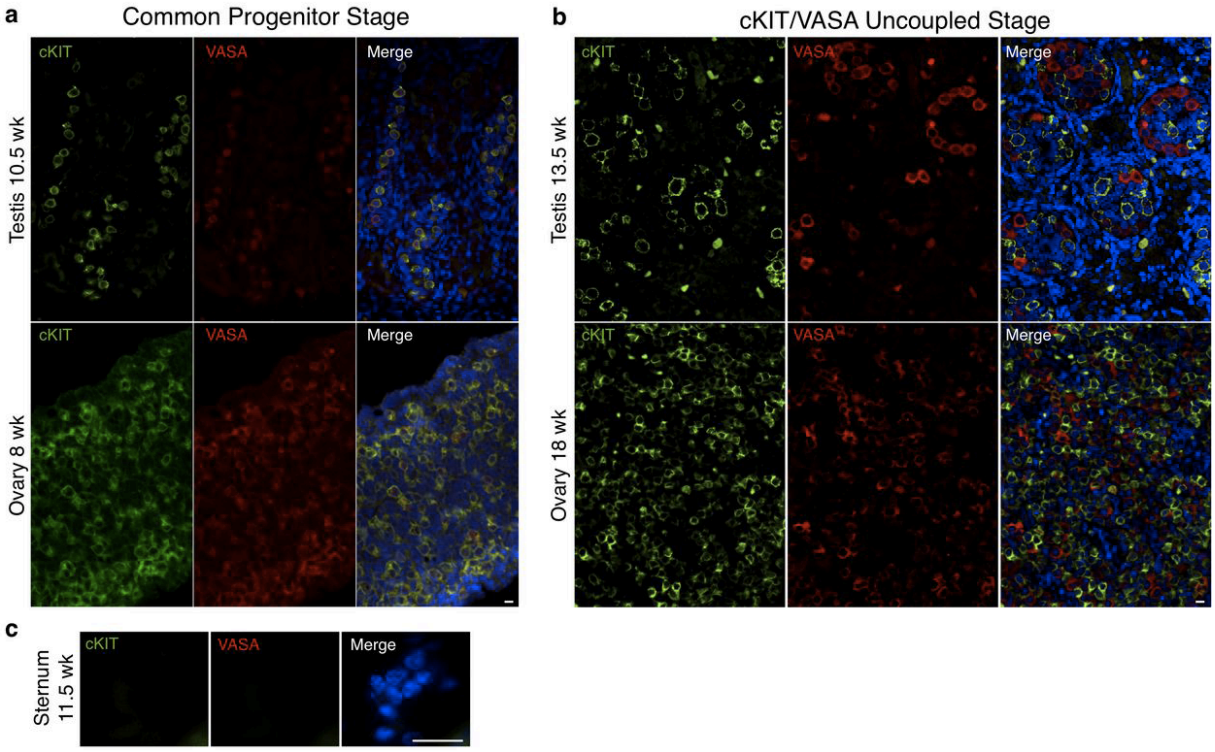

Figure 4-S2. cKIT and VASA expression in fetal testes and ovaries.

Representative immunofluorescence images of cKIT with VASA, (a) at the common progenitor stage, in a 10.5-week testes and an 8-wk ovary, (b) at the cKIT/VASA uncoupled stage, in a 13.5-week testis and an 18-week ovary. (c) Immunofluorescence of cKIT with VASA on 11.5-week sternum tissue as negative control for cKIT and VASA antibody specificity on paraffin sections. Nuclei were counterstained with DAPI (blue). Scale bars represent 10 um. Abbreviations: wk= week.

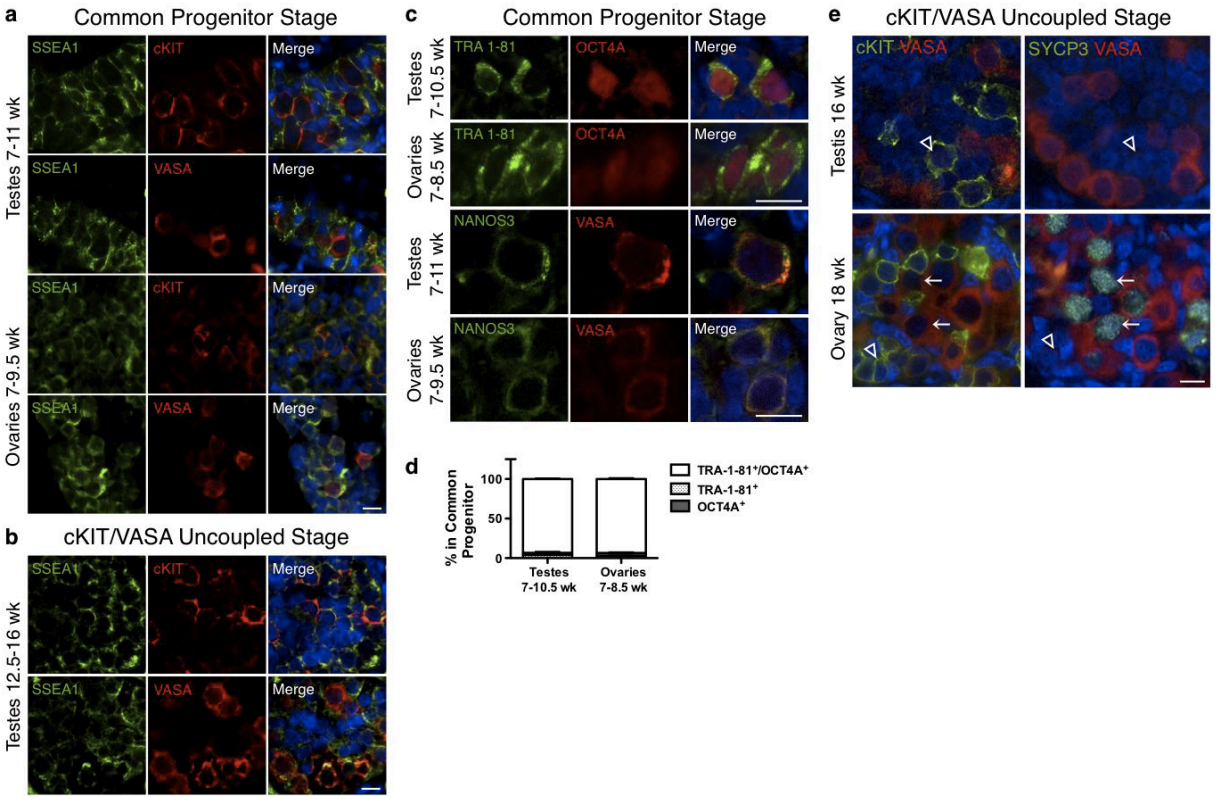

Figure 4-S3. Dynamics of cKIT, OCT4 and/or VASA with SSEA1, TRA-1-81, NANOS3 and SYCP3.

(a,b) Representative immunofluorescence images of SSEA1 with cKIT or VASA, (a) at the common PGC progenitor stage, in testes shown is an 11-week sample for 7-11wk (n=2) and an 8-week ovary for 7-9.5wk (n=2). (b) The cKIT/VASA uncoupled stage in the testis from 12.5-16wk (n=3) shown is a testes at 13.5-weeks. (c) Representative immunofluorescence images of TRA-1-81 with OCT4A and NANOS3 with VASA at the common PGC progenitor stage. For TRA-1-81, shown is a 10.5-week testis for 7-10.5wk (n=3) and an 8.5-week ovary for 7-8.5wk (n=3). For NANOS3 shown is an 11-week in testis for 7-11 wk (n=3) and a 9.5-week in ovary for 7-9.5wk (n=2). (d) Quantification of TRA-1-81^{+}, OCT4A^{+} and TRA-1-81^{+}/OCT4A^{+} in testes at 7-10.5wk, 6 optic fields counted (n=3) and in ovaries at 7-8.5wk, 7 optic fields counted (n=3). (e) Representative immunofluorescence images at the cKIT/VASA uncoupled stage of VASA with cKIT or SYCP3 on adjacent sections at 16 weeks in testis and 18 weeks in ovary. Arrows indicate the same cKIT^{+}, and open arrowheads the same VASA^{+} cell in adjacent sections. Nuclei were counterstained with DAPI (blue). Scale bars represent 10 um. All data are mean ± SEM.

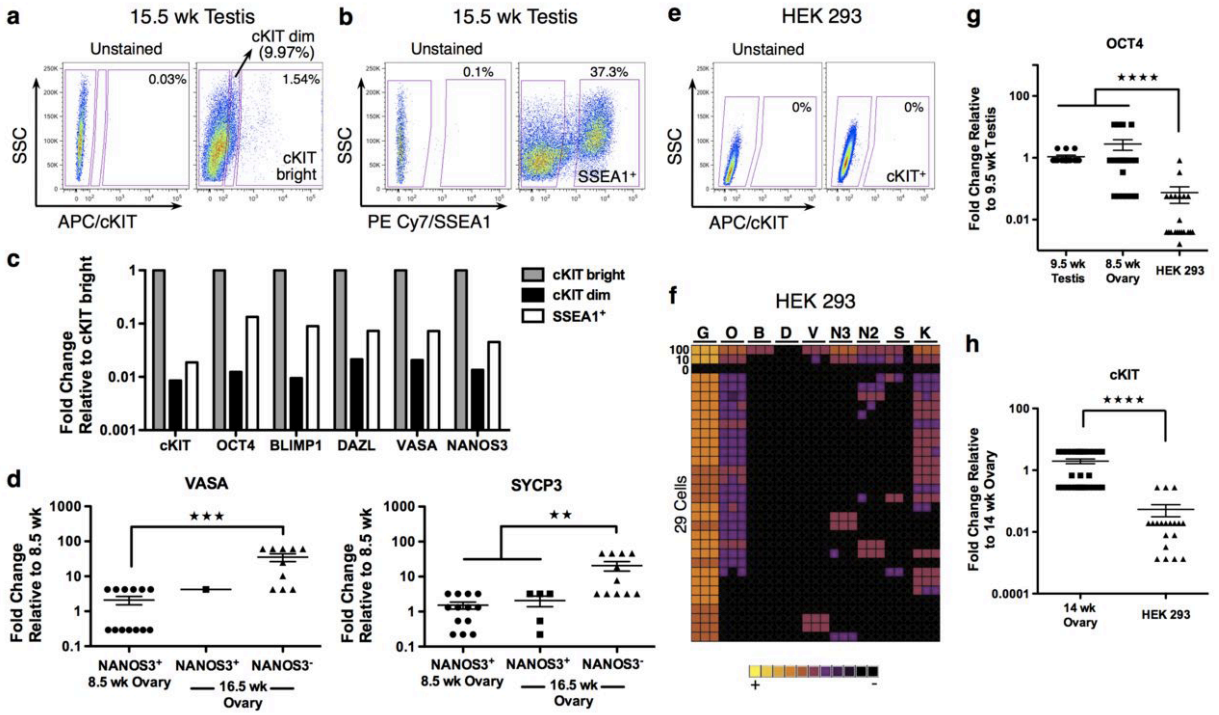

Figure 4-S4. Germ line identity is enriched in the cKIT bright fraction in 15.5-week testis.

(a,b) Gating strategy for sorting, (a) cKIT dim and cKIT bright cells with an APC conjugated anti-human cKIT primary antibody and (b) SSEA1$^+$ cells using PE Cy7 secondary antibody against side scatter (SSC) on a 15.5-week testis. Indicated above the graph is the percent positive cells sorted in each gate (c) qRT-PCR showing fold change enrichment in cKIT dim and SSEA1$^+$ relative to the cKIT bright fraction for *cKIT*, *OCT4*, *BLIMP1*, *DAZL*, *VASA* and *NANOS3* after *GAPDH* normalization. Shown on a log10 scale. (d) qRT-PCR data analysis for *VASA* and *SYCP3* in *NANOS3* expressing cKIT$^+$ cells sorted from the ovary at 8.5 weeks, and *NANOS3* positive or negative cKIT$^+$ cells sorted from the ovary at 16.5 weeks. Levels are normalized to *GAPDH* and are shown as fold change relative to 8.5 weeks on a log10 scale. Each data point represents n=1 single cell. (e) Flow plot of HEK 293 stained with an APC conjugated anti-human cKIT primary antibody against side scatter (SSC), showing there are no cKIT expressing cells. (f) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3), *NANOS2* (N2), *SYCP3* (S) and *cKIT* (K) in triplicate (columns) in 100, 10, 0 or single cells (rows) in sorted HEK 293 cells. Note that HEK 293 cells have previously been reported to express *OCT4* mRNA[15] and we show that cKIT is transcribed but the protein is not expressed on the cell surface (g) qRT-PCR analysis for *OCT4* in single cKIT$^+$ cells sorted from the fetal testes at 9.5 weeks, the fetal ovary at 8.5 weeks, and sorted HEK 293 cells. (h) qRT-PCR data analysis for *cKIT* in cKIT$^+$ cells sorted from the fetal ovary at 14 weeks and sorted HEK 293s. Levels are normalized to *GAPDH* and are shown as fold change referenced to the 9.5 wk testis in (g) and the 14 wk ovary in (h) on a log10 scale. Note that where individual HEK 293 cells express *OCT4* and *cKIT* the levels are 10-to-100 fold lower than the reference germ line cells. Each data point represents n=1 cell. All data are mean ± SEM. (*P* values were determined using a Mann-Whitney test, ⋆⋆ p<0.005, ⋆⋆⋆ p<0.0005 and ⋆⋆⋆⋆ p<0.0001).

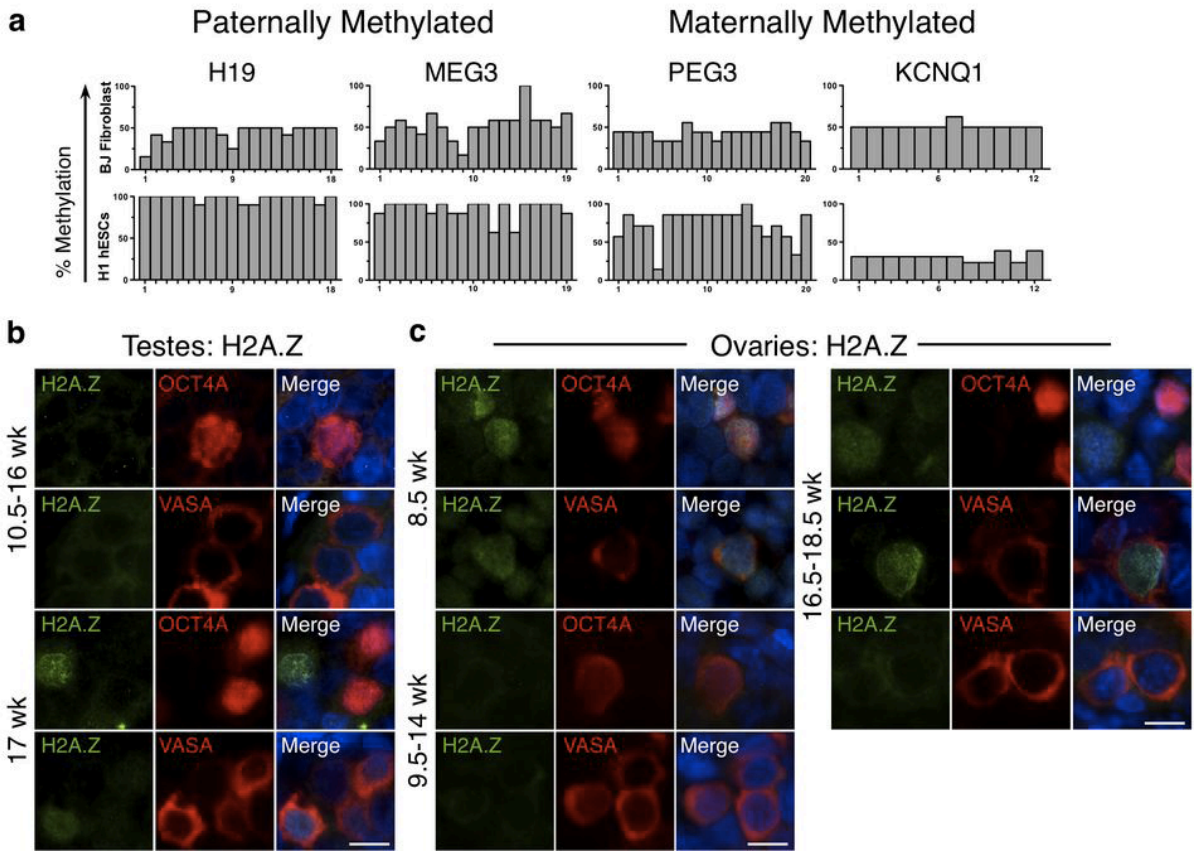

Figure 4-S5.

(a) Methylation Analysis of BJ fibroblast line and H1 hESC line performed by BS-PCR on *H19*, *MEG3*, *PEG3* and *KCNQ1* DMRs in BJ fibroblast and H1 hESCs. (b,c) Fetal PGCs in the testis and ovary become transiently devoid of H2A.Z. Representative immunofluorescence images of H2A.Z with OCT4A or VASA in (b) testes and (c) ovaries at designated developmental ages. (b) Shown is a 13.5-week testis for 10.5-16wk (n=5) and a 17-week testis (n=1). (c) Shown is an 8.5-week ovary (n=2), a 14-week ovary for 9.5-14wk (n=3) and a 18.5-week ovary for 16.5-18.5wk (n=3). Abbreviations, wk=week. Nuclei were counterstained with DAPI (blue). Scale bars represent 10 um.

REFERENCES

1. Ohinata Y. *et a*l. A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137,** 571-84 (2009).

2. Hayashi K. *et al*. Offspring from Oocytes Derived from in Vitro Primordial Germ Cell-Like Cells in Mice. *Science* 2012 Oct 4. [Epub ahead of print]

3. Reijo R. *et al*. Mouse autosomal homolog of DAZ, a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty. *Genomics* **35**, 346-352 (1996).

4. Park, T. S. *et al.* Derivation of primordial germ cells from human embryonic and induced pluripotent stem cells is significantly improved by co-culture with human fetal gonadal cells. *Stem Cells* **27**, 783–795 (2009).

5. Gaskell, T. L., Esnal, A., Robinson, L. L. L., Anderson, R. A. & Saunders, P. T. K. Immunohistochemical profiling of germ cells within the human fetal testis: identification of three subpopulations. *Biol Reprod* **71**, 2012–2021 (2004).

6. Robinson, L., Gaskell, T., Saunders, P. & Anderson, R. Germ cell specific expression of c-kit in the human fetal gonad. *Molecular human reproduction* **7**, 845 (2001).

7. Høyer, P. E., Byskov, A. G. & Møllgård, K. Stem cell factor and c-Kit in human primordial germ cells and fetal ovaries. *Molecular and cellular endocrinology* **234**, 1–10 (2005).

8. Pauls, K. Spatial expression of germ cell markers during maturation of human fetal male gonads: an immunohistochemical study. *Hum Reprod* **21**, 397–404 (2005).

9. Kerr, C., Hill, C., Blumenthal, P. & Gearhart, J. Expression of pluripotent stem cell markers in the human fetal ovary. *Human reproduction* **23**, 589 (2008).

10. Kerr, C. L., Hill, C. M., Blumenthal, P. D. & Gearhart, J. D. Expression of pluripotent stem cell markers in the human fetal testis. *Stem Cells* **26**, 412–421 (2008).

11.Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**, 440–458 (2005).

12. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877–881 (2008).

13. Hajkova, P. *et al.* Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. *Science* **329**, 78–82 (2010).

14. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).

15. Sinclair, A.H. *et al*. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240–244 (1990).

16. Atlasi, Y., Mowla, S. J., Ziaee, S. A. M., Gokhale, P. J. & Andrews, P. W. OCT4 Spliced Variants Are Differentially Expressed in Human Pluripotent and Nonpluripotent Cells. *Stem Cells* **26**, 3068–3074 (2008).

17. Warthemann, R., Eildermann, K., Debowski, K. & Behr, R. False-positive antibody signals for the pluripotency factor OCT4A (POU5F1) in testis-derived cells may lead to erroneous data and misinterpretations. *Molecular human reproduction* (2012).doi:10.1093/molehr/gas032

18. Ruzov, A. *et al.* Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res* **21**, 1332–1342 (2011).

19. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010).

20. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).

21.Koh, K. P. *et al.* Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells. *Stem Cell* **8**, 200–213 (2011).

22. Thomson, J. A. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282**, 1145–1147 (1998).

23. Diaz Perez, S. V. *et al.* Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. *Human Molecular Genetics* **21**, 751–764 (2012).

24. Clark, A. T. Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. *Human Molecular Genetics* **13**, 727–739 (2004).

25. Zwaka, T. P. A germ cell origin of embryonic stem cells? *Development* **132**, 227–233 (2005).

26. Bao, S. *et al.* The Germ Cell Determinant Blimp1 Is Not Required for Derivation of Pluripotent Stem Cells. *Cell Stem Cell* **11**, 110–117 (2012).

27. Kamei, K. *et al.* Microfluidic image cytometry for quantitative single-cell profiling of human pluripotent stem cells in chemically defined conditions. *Lab Chip* **10**, 1113 (2010).

28. Briddell, R. *et al.* Further Phenotypic Characterization and Isolation of Human Hematopoietic Progenitor Cells Using a Monoclonal-Antibody to the C-Kit Receptor. *Blood* **79**, 3159–3167 (1992).

29. Anderson, R., Fulton, N., Cowan, G., Coutts, S. & Saunders, P. Conserved and divergent patterns of expression of DAZL, VASA and OCT 4 in the germ cells of the human fetal ovary and testis. *BMC Dev Biol* **7**, 136 (2007).

30. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).

31. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential implications for active demethylation of CpG sites. *J Biol Chem* **286**, 35334–35338 (2011).

32. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15–23 (2002).

33. Ohinata, Y., Sano, M., Shigeta, M., Yamanaka, K. & Saitou, M. A comprehensive, non-invasive visualization of primordial germ cell development in mice by the Prdm1-mVenus and Dppa3-ECFP double transgenic reporter. *Reproduction* **136**, 503–514 (2008).

34. Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* **22**, 1617–1635 (2008).

35. Kee, K., Gonsalves, J. M., Clark, A. T. & Pera, R. A. R. Bone morphogenetic proteins induce germ cell differentiation from human embryonic stem cells. *Stem Cells Dev* **15**, 831–837 (2006).

36. Kee, K., Angeles, V. T., Flores, M., Nguyen, H. N. & Pera, R. A. R. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature* **462**, 222–225 (2009).

37. Aflatoonian, B. *et al.* In vitro post-meiotic germ cell development from human embryonic stem cells. *Hum Reprod* **24**, 3150–3159 (2009).

38. Tilgner, K. *et al.* Expression of GFP under the control of the RNA helicase VASA permits fluorescence-activated cell sorting isolation of human primordial germ cells. *Stem Cells* **28**, 84–92 (2010).

39. Panula, S. *et al.* Human germ cell differentiation from fetal-and adult-derived induced pluripotent stem cells. *Human Molecular Genetics* **20**, 752–762 (2011).

40. Medrano, J. V., Ramathal, C., Nguyen, H. N., Simon, C. & Reijo Pera, R. A. Divergent RNA-binding Proteins, DAZL and VASA, Induce Meiotic Progression in Human Germ Cells Derived in Vitro. *Stem Cells* **30**, 441–451 (2012).

41. Chuang, C. Y. *et al.* Meiotic Competent Human Germ Cell-like Cells Derived from Human Embryonic Stem Cells Induced by BMP4/WNT3A Signaling and OCT4/EpCAM (Epithelial Cell Adhesion Molecule) Selection. *J Biol Chem* **287**, 14389–14401 (2012).

42. Plath, K. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science* **300**, 131–135 (2003).

43. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).

44. Vincent, J. J. *et al.* Single Cell Analysis Facilitates Staging of Blimp1-Dependent Primordial Germ Cells Derived from Mouse Embryonic Stem Cells. *PLoS ONE* **6**, e28960 (2011).

45. Boissonnas, C. *et al.* Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *European Journal of Human Genetics* **18**, 73–80 (2009).

46. Geuns, E., Hilven, P., Van Steirteghem, A., Liebaers, I. & De Rycke, M. Methylation analysis of KvDMR1 in human oocytes. *Journal of Medical Genetics* **44**, 144–147 (2006).

47. Kagami, M. *et al.* The IG-DMR and the MEG3-DMR at Human Chromosome 14q32.2: Hierarchical Interaction and Distinct Functional Properties as Imprinting Control Centers. *PLoS Genet* **6**, e1000992 (2010).

48. Zechner, U. *et al.* Quantitative methylation analysis of developmentally important genes in human pregnancy losses after ART and spontaneous conception. *Molecular human reproduction* **16**, 704–713 (2010).

CHAPTER 5

CONCLUSIONS

The primary goal of the studies presented in this dissertation is to identify and characterize the mechanisms that drive the earliest stages of germ line establishment, development, and epigenetic remodeling. In Chapter 2, we developed the means to differentiate and isolate PGC-like in vitro derivatives from ESCs that faithfully model the early pre-gonadal stages of development, and successfully employed single cell gene expression technology to profile developmental staging of in vitro derived PGCs. This technology enabled our ability to genetically manipulate the development of PGC in vitro and, and led to the discovery that 5mC modifying enzymes do not mediate the genome-wide debulking of methylated cytosine bases from DNA in Chapter 3. We then successfully employed our single cell gene expression approach to understand key developmental and epigenetic events in the human germ line using cKIT as a sortable marker of human PGCs from gonads in Chapter 4. Together, this work establishes a technology and differentiation methodology to characterize early events in germ line development, and a means to apply these techniques and principles to further characterization of germ line development in vitro.

The development of the iPGC differentiation system is a significant advancement over other previously established models of germ line formation from ES cells. Transgene-free methodologies for generating high-purity populations of iPGCs have only recently been developed by our group as well as others [1-3]. Using this system to study the gonad-independent phase of PGC development facilitates easy genetic manipulation of ESCs to screen for pathways and genes involved in PGC development. This was proven by proof-of-principle in Chapter 2 by deletion of *Blimp1,* and was critical in the depletion of Tet proteins in Chapter 3. This system is therefore compatible with pre-existing tools, including mutant ESC lines and viral to manipulate gene expression. Finally, the scalability of this system is a huge advancement over all other known PGC differentiation methods, which generally yield very few bona fide

PGCs, or a mixed population of PGC progenitors. This advantage has lead to the first-ever genome-wide profiling map of methylation by BS-Seq of PGCs in vitro, and certainly the first to characterize Step 1 DNA demethylation at the sequence level. These advantages make this system highly adaptable and amenable to further examination of the chromatin of early PGCs, such as histone modifications that are thought to repress somatic gene expression and hypothesized to be regulated by histone tail arginine methylation [4,5].

It is important to note that floating culture differentiation of PGCs has consistently resulted in the generation of iPGCs with distinct pre-gonadal features, including lack of Mvh expression and maintenance of imprint methylation. Indeed, PGC maturation techniques leading to functional gametes has required contact with a gonadal environment to produce maturation and ultimately spermatogenesis or oogenesis *in vivo* via transplantation [1,2]. Therefore, the utility of the technology presented here allows for the study of Step 1 reprogramming DNA methylation—and perhaps other epigenetic features of Step 1 PGCs—uncoupled from epigenetic and developmental events associated with gonad colonization.

Single cell gene expression techniques were critical in our establishment of an in vitro model of PGC differentiation, and were useful in dissecting distinctions between early pluripotent cells and bona fide PGCs from embryonic stem cells. Given the shared expression of pluripotency-associated genes, as well as low to sporadic expression of PGC markers within undifferentiated ESC cultures [3], the ability to address PGC identity at a single cell level has been impaired. Prior to the studies of Chapter 2, the only test of PGC identity relied upon the ability of putative PGCs to self-renew in the presence of retinoic acid [6]. Our experiments clearly demonstrated å

stochastic pattern of PGC genes in undifferentiated ESCs, and in all cases were the levels of these genes significantly lower than in endogenous or iPGCs (Chapter 2).

Our in vitro studies have concluded that Tet1 and Tet2 do not play a role in the genome wide loss of 5mC from the genome of iPGCs during EB differentiation. In future studies, the dynamics of 5hmC in PGCs during the specification and pre-reprogramming period (E7.25-E8.5) should be examined *in situ*. While our studies suggest that 5hmC is not a global mediator of demethylation, we can speculate that 5hmC and Tet proteins have locus-specific roles in the establishment of PGC epigenomes prior to the onset of migration. This could be addressed via immunofluorescence for 5hmC to determine 5hmC content prior to reprogramming, as well as lineage-trace sorting of newly specified PGCs (for example, *Dppa3-Cre x R26R-GFP* embryos). Our finding that loss of *Tet1* affects the methylation of meiotic loci in iPGCs indicates that regulation of meiotic genes is at the methylation level, is cell-intrinsic, and occurs prior to gonadal colonization. These results implicate DNA demethylation as a regulator of not only totipotency, but of later germ line functionality, perhaps as a developmental timer to repress premature meiotic gene expression.

Towards the Clinic: Germ Cell Replacement Therapy?

One of the most attractive clinical applications for cell replacement therapy is the concept of generating patient-specific gametes from induced pluripotent stem cells (iPSCs) derived from a somatic sample of a patient [7,8]. This would theoretically be an attractive option for sterile men, or for men who have undergone gonadotoxic therapy, a common result of cancer treatment [9]. PGCs have successfully been generated from human ESCs as well as iPSCs, providing promise for this approach [10-12], (Chapter 4).

However, recent studies have shown a requirement for the somatic cellular microenvironment for gametogenesis from in vitro derived germ cells. In mice, PGCs derived from ESCs and iPGCs give rise to mature germ cells capable of generating viable that generate fertile offspring [1,2,13]. In these studies PGC-like cells were only functional after transplantation in host gonadal environments or reconstructed environments. The development of this technology is a significant advance that can be adapted to assay for the consequences of genetic manipulation of PGCs during early phases on gametogenesis and fertility. These results are also promising proof-of-principle experimental approaches for the possibility of a human therapeutic approach, however this would require the development of transplantation approaches in primate models, particularly in determining functionality of human ESC-derived germ cells. Fortunately, recent primate advancements in transplantation of spermatogonial stem cells (SSCs) may be adaptable to address these questions [14].

In Chapter 3, we sought to understand the means of DNA demethylation in PGCs prior to gonadal colonization. Understanding DNA demethylation in PGCs has therapeutic implications for therapeutic reprogramming, and for understanding the epigenetic basis of totipotency. Acquisition of DNA methylation and other epigenetic marks is thought to establish or maintain cellular identity [15-18], and methylation patterns are relatively stable after they are established [19]. DNA demethylation of somatic gene loci does occur during somatic cell reprogramming, where iPSCs undergo promoter-specific demethylation of pluripotency associated genes [7,8]. Partially reprogrammed iPSCs which proliferate in culture but do not fully re-activate endogenous pluripotency genes show DNA hypermethylation of pluripotency-associated genes, indicating it is a bottleneck in the reprogramming process [20]. Active DNA demethylation during

reprogramming via heterokaryon formation is associated with achieving the fully reprogrammed state ([21]. Nonetheless, pluripotent iPSCs derived from somatic cells harbor a DNA methylation signature reminiscent of the donor cell-type, resulting in pre-disposed differentiation potential toward that cell type [22-25]. Indeed, addition of chromatin-modifying drugs was able to remove the somatic epigenetic memory of iPSCs, indicating that epigenetic recodification of the somatic genome is resistant to epigenetic remodeling, despite activation of endogenous pluripotent transcriptional pathways and circuit [25]. These defects are of utmost importance when considering iPSCs as a cell source for germ cells, which must be able to convey epigenetic totipotency during embryo fertilization. Therefore the pathways that underlie these processes are the role of epigenetics in genome inheritance, and crucial to approaches to experimentally define epigenetic landscapes for targeted differentiation strategies.

REFERENCES

1 Hayashi, K. *et al.* Offspring from Oocytes Derived from in Vitro Primordial Germ Cell-Like Cells in Mice. *Science*, doi:10.1126/science.1226889 (2012).

2 Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532, doi:10.1016/j.cell.2011.06.052 (2011).

3 Vincent, J. J. *et al.* Single cell analysis facilitates staging of Blimp1-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS One* **6**, e28960, doi:10.1371/journal.pone.0028960 (2011).

4 Ancelin, K. *et al.* Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat Cell Biol* **8**, 623-630, doi:10.1038/ncb1413 (2006).

5 Tee, W. W. *et al.* Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes Dev* **24**, 2772-2777, doi:10.1101/gad.606110 (2010).

6 Geijsen, N. *et al.* Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* **427**, 148-154, doi:10.1038/nature02247 (2004).

7 Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141-146, doi:10.1038/nature06534 (2008).

8 Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872, doi:10.1016/j.cell.2007.11.019 (2007).

9 Clark, A. T., Phillips, B. T. & Orwig, K. E. Fruitful progress to fertility: male fertility in the test tube. *Nature medicine* **17**, 1564-1565, doi:10.1038/nm.2594 (2011).

10 Clark, A. T. *et al.* Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. *Hum Mol Genet* **13**, 727-739, doi:10.1093/hmg/ddh088 (2004).

11 Kee, K., Angeles, V. T., Flores, M., Nguyen, H. N. & Pera, R. A. R. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature*, 1-6, doi:10.1038/nature08562 (2009).

12 Park, T. S. *et al.* Derivation of Primordial Germ Cells from Human Embryonic and Induced Pluripotent Stem Cells Is Significantly Improved by Coculture with Human Fetal Gonadal Cells. *Stem Cells* **27**, 783-795, doi:10.1002/stem.13 (2009).

13 Ohinata, Y. *et al.* A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. *Cell* **137**, 571-584, doi:10.1016/j.cell.2009.03.014 (2009).

14 Hermann, B. P. *et al.* Spermatogonial stem cell transplantation into rhesus testes regenerates spermatogenesis producing functional sperm. *Cell stem cell* **11**, 715-726, doi:10.1016/j.stem.2012.07.017 (2012).

15 Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330**, 612-616, doi:10.1126/science.1191078 (2010).

16 Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).

17 Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139**, 15-31, doi:10.1242/dev.050849 (2012).

18 Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews. Genetics* **9**, 465-476, doi:10.1038/nrg2341 (2008).

19 Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8689-8694, doi:10.1073/pnas.1002720107 (2010).

20 Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49-55, doi:10.1038/nature07056 (2008).

21 Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042-1047, doi:10.1038/nature08752 (2010).

22 Bar-Nur, O., Russ, H. A., Efrat, S. & Benvenisty, N. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell stem cell* **9**, 17-23, doi:10.1016/j.stem.2011.06.007 (2011).

23 Kim, K. *et al.* Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nature biotechnology* **29**, 1117-1119, doi:10.1038/nbt.2052 (2011).

24 Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat Cell Biol* **13**, 541-549, doi:10.1038/ncb2239 (2011).

25 Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285-290, doi:10.1038/nature09342 (2010).