

UCLA

UCLA Electronic Theses and Dissertations

Title

Naive Pluripotency in Human Primordial Germ Cell Development

Permalink

<https://escholarship.org/uc/item/9bg21874>

Author

Hancock, Grace

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Naive Pluripotency in Human Primordial Germ Cell Development

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of
Philosophy in Molecular Biology

by

Grace VerHage Hancock

2021

© Copyright by
Grace VerHage Hancock
2021

ABSTRACT OF THE DISSERTATION

Naive pluripotency in Human Primordial Germ Cell Development

by

Grace VerHage Hancock

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2021

Professor Amander Therese Clark, Chair

Infertility is caused by a multitude of genetic and environmental factors, and many who seek treatment are unresponsive to *in vitro* fertilization. Treatments such as these are costly, as well as physically and emotionally demanding, so it is critical to better understand the individual underlying causes of infertility to improve patient care. Germ cells give rise to the next generation and are responsible for passing along genetic and epigenetic information, and abnormalities in germ cell development can result in infertility. In addition, mutations in certain genes related to germ cell formation can contribute to their improper development, resulting in germ cell tumor formation or other developmental disorders in their future offspring. In these studies, we investigated the earliest cell type in the human germline, known as human Primordial Germ Cells (hPGCs). We developed models and techniques to represent hPGC formation and identified genes of interest to investigate their potential role in healthy hPGC identify and development.

Two genes, KLF4 and TFCEP2L1, are upregulated in hPGCs and are known to have a role in naive pluripotency. Given this, and other shared characteristics between the hPGCs and naive pluripotency, we hypothesized that KLF4 and TFCEP2L1 have a role in hPGC development. We

first used CRISPR/Cas9 gene editing technology at the human embryonic stem cell (hESC) stage of our PGC-like cell (hPGCLC) aggregate model, to demonstrate the requirement of certain genes including EOMES in PGCLC induction. Then, we developed an extended culture system to study hPGCLCs *in vitro*. In extended culture, hPGCLCs begin the process of reprogramming to represent the later hPGC stage, where we can study characteristics such as cell cycle and survival. We knocked out KLF4 and TFCEP2L1 separately in hESC lines, and discovered that neither gene is required for PGCLC induction, but each has an anticipated role in re-acquiring naive pluripotency in the hESC state. Despite these functional null mutations, neither gene is required for progression of hPGCLC development *in vitro* including in progression through S phase as measured by Edu, or for PGCLC colony survival. Given this, we conclude that the naive pluripotent states observed in the pre-implantation and hPGC stages are uniquely controlled, and that these transcription factors may serve alternate roles in hPGC development. Possible roles include protection against germ cell tumors, or facilitation of later and advanced stage hPGC development. Further work will address these possibilities, using animal and stem cell models of gonadal development to study later stage progression.

The dissertation of Grace Hancock is approved.

Patrick Allard

Jeffery Aaron Long

Karen Marie Lyons

Kathrin Plath

Amander Therese Clark, Committee Chair

University of California, Los Angeles

2021

Dedication

To Taylor Brown, who never failed to share her smile.

Table of Contents

ABSTRACT OF THE DISSERTATION.....	ii
Committee Page.....	iv
Dedication.....	v
Table of Contents.....	vi
List of Figures.....	viii
List of Tables.....	viii
Acknowledgements.....	ix
VITA.....	xiv
Chapter 1 – Introduction	1
Impact of the germline in human health.....	2
hPGC specification.....	3
Models of late-stage PGC development.....	5
Pluripotency.....	5
References.....	7
Chapter 2 – Mammalian Primordial Germ Cell Specification.....	10
Abstract.....	11
Introduction.....	11
PGC specification in the mouse embryo	11
PGC specification in the monkey embryo	12
PGC specification in the human embryo.....	15
DNA methylation remodeling and X-chromosome reactivation in PGCs.....	15
<i>In vitro</i> and <i>Ex vivo</i> models of PGC specification.....	16
Conclusions and future perspectives.....	19
References.....	19
Chapter 3 – Germline competency of human embryonic stem cells depends on Eomesodermin.....	23
Abstract.....	24
Introduction.....	25
Material and methods.....	25
Results	28
Discussion.....	32
References.....	34

Chapter 4 – An extended culture system that supports human primordial germ cell-like survival and initiation of DNA Methylation Erasure.....	36
Summary.....	37
Introduction.....	37
Results.....	38
Discussion.....	45
Experimental procedures.....	46
Reference.....	48
Chapter 5 – Divergent roles for KLF4 and TFCP2L1 in naive ground state pluripotency and human primordial germ cell development.....	51
Summary.....	53
Introduction.....	53
Results.....	56
Discussion.....	61
Experimental procedures.....	63
Acknowledgements.....	72
Author Contributions.....	73
References.....	74
Figures and legends.....	79
Supplementary Tables	88
Chapter 6 – Conclusion	89
References.....	93

List of Figures

Figure 1-1: Developing stable hESC cell lines with CRISPR/Cas9 gene editing and colony screening.....	4
Figure 2-1: Representation of peri-implantation embryo development in mammals.....	13
Figure 2-2: Stem cell models of primordial germ cell induction.....	17
Figure 3-1: Analysis of ITGA6/EPCAM and TNAP/cKIT populations in human prenatal gonads.....	29
Figure 3-2: Germline competency varies between independent hESC lines.....	31
Figure 3-3: TGFb and WNT signaling are required for PGCLC induction from hESCs.....	32
Figure 3-4: EOMES is required for PGCLC induction from hESCs.....	33
Figure 4-1: hPGCLCs cultured on STOs maintain germline identity.....	39
Figure 4-2: hPGCLCs in extended culture maintain a transcriptome similar to specified hPGCLCs.....	42
Figure 4-3: hPGCLCs in extended culture self-renew and undergo expansion.....	43
Figure 4-4: Partial chromatin remodeling is maintained in extended culture hPGCLCs.....	44
Figure 4-5: hPGCLCs in extended culture undergo partial genome-wide demethylation.....	45
Figure 5-1: KLF4 and TFCEP2L1 are dynamically expressed during hPGCLC specification.....	79
Figure 5-2: KLF4 and TFCEP2L1 are required for naive pluripotency in 5iLAF.....	80
Figure 5-3: KLF4 and TFCEP2L1 are not involved in hPGCLC induction from hESCs.....	81
Figure 5-4: hPGCLCs survive in extended culture and express TFCEP2L1.....	82
Figure 5-5: KLF4 and TFCEP2L1 are not required for proliferation in extended culture.....	83
Figure 5-6: KLF4 and TFCEP2L1 are dynamically regulated in the germline.....	84
Figure 5-7: KLF4 and TFCEP2L1 mutant hESC subline characterization.....	85
Figure 5-8: KLF4 knockout does not affect the global transcriptome.....	86
Figure 5-9: KLF4 and TFCEP2L1 do not regulate hPGCLC survival.....	87

List of Tables

Table 2-1: Key features of primordial germ cell (PGC) development in different mammals.....	14
Table 4-1: Key components of 7F and FR10 media.....	40
Table 5-1: Guide RNA and primer sequences for CRISPR mutations.....	88
Table 5-2: RNA sequencing samples and mapping.....	88

Acknowledgements

This work would not have been possible without the multitude of individuals who supported me throughout my PhD. First I would like to thank my thesis advisor, Dr. Amander Clark, for guiding my research in the right direction, and always pushing me forward. She was always empathetic to how difficult research is, and never let me give up even when I wanted to. I will carry these valuable lessons about resilience with me well beyond the lab. The environment she has cultivated in our laboratory is cooperative, friendly, exciting, and supportive. I would not have wanted to work with any other group.

I would like to thank all of the lab members in the Clark lab, many who contributed to these research findings, and all who contributed to making my graduate experience as enjoyable as possible. Thank you to Dr. Di Chen for all of his patient and fun mentorship when I was just getting started and had so much to learn. To my undergraduate mentee, Lior Peretz for working alongside me throughout so many difficulties of this project, never losing her cool when challenges arose. To Dr. Sissy Wamaitha for her friendship inside and outside of the lab, sharing her knowledge of development and how to survive a PhD working on stem cells. Thank you to the current members and lab alumni who were there from the beginning and supported my growth: Timothy Hunt, Dr. Enrique Sosa, Dr. Joanna Gell, Dr. Yu Tao. To Matthew Lowe and Tsotne Chitiashvili for sharing the work with me and chatting about anything and everything over the years. Thank you to Dr. Wanlu Liu for her patient and generous help with data analysis, her friendship, and for being a role model to me. To current lab members who were so pleasant to work with during this last year of the pandemic and who are moving the lab in exciting directions: Dr. Erica Pandolfi, Dr. Fei-man Hsu, and Jon DiRusso. Thank you to the past and

current undergraduates for all your help and energy: Ernesto Rojas, Jared Faith, Esme Villavicencio, Marianna Aslanyan, Jill Zimmerman, Yao Chang Tan, Grace Bower, Isaac Gorgy, Allison Wang, and Amanda Hagen. Finally, thank you to everyone who fed my cells at any point—the real heroes of a stem cell group.

I would like to thank my committee members, Dr. Patrick Allard, Dr. Jeffery Long, Dr. Karen Lyons, and Dr. Kathrin Plath for their advice, direction, and support since 2016. As my project changed and hit roadblocks they remained supportive while allowing me to grow as a scientist. Additionally, thank you to Jeff Long for being a great interim committee chair and the best home area director for the best home area. Thank you to our student affairs officers, Ashley TerHorst and Stephanie Cuellar for always having students' best interest in mind. To Diana Azurdia for her empathy and vulnerability when I needed it the most. Thank you to Jessica Mcgruder, Felicia Codrea, and Jefferey Camlimlim at the flow cytometry core for consistently helping me produce beautiful data, and for their genuine care and interest in my day to day life. Thank you very much to the UCLA staff, facilities, and maintenance workers who work so hard to keep our research running without interruption, especially during this year's pandemic and shutdown.

Thank you to my family, for supporting me in many ways from near and far. For allowing me the space to tackle this on my own and grow as an individual, and always understanding my challenging schedule.

Finally, thank you to my graduate student classmates and friends. Whether we met years ago or this week you have made this experience worth it. Celebrating and commiserating with you has been invaluable. I appreciate all your hard work and am honored to have met and shared this with you.

Chapter 2 is a review article originally published in *Development*. Hancock, G. V., Wamaitha, S.E., Peretz, L., & Clark, A. T. (2021). Mammalian primordial germ cell specification. *Development*, 148(dev189217). <https://doi.org/10.1242/dev.189217>. It is reprinted here with permission. The authors' research was funded by grants from the National Institutes of Health/Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD079546, R01HD098278 and R01HD058047). S.W. is supported by funds from the Iris Cantor-UCLA Women's Health Center Executive Advisory Board (NCATS UCLA CTS1 grant number UL1TR001881).

Chapter 3 was originally published in *Biology of Reproduction*. Chen, D., Liu, W., Lukianchikov, A., Hancock, G. V., Zimmerman, J., Lowe, M. G., Kim, R., Galic, Z., Irie, N., Surani, M. A., Jacobsen, S. E., & Clark, A. T. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. *Biology of Reproduction*, 97(6), 850–861. <https://doi.org/10.1093/biolre/iox138>. It is reprinted here with permission. This work is supported by the Medical Research Council ([MR/P009948/1/MRC](https://doi.org/10.1093/biolre/iox138)), and NIH ([R01 HD079546/HD/NICHD, R24 HD000836/HD/NICHD](https://doi.org/10.1093/biolre/iox138)). The authors would like to thank Felicia Codrea and Jessica Scholes for FACS, Jinghua Tang for banking and culturing of the UCLA hESC lines, 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. Human conceptus tissue requests can be made to bdrl@u.washington.edu. SEJ is an investigator of the Howard Hughes Medical Institute.

Chapter 4 was originally published in Stem Cell Reports. Gell, J. J., Liu, W., Sosa, E., Chialastri, A., Hancock, G., Tao, Y., Wamaitha, S. E., Bower, G., Dey, S. S., & Clark, A. T. (2020). An Extended Culture System that Supports Human Primordial Germ Cell-like Cell Survival and Initiation of DNA Methylation Erasure. *Stem Cell Reports*.

<https://doi.org/10.1016/j.stemcr.2020.01.009>. It is reprinted with permission. This project was funded with supported from the Pablove Foundation 20183715 (to J.J.G.), R01HD079546 (to A.T.C.), and supplement to R01 R01HD058047 (to A.T.C.), and UC Cancer Research Coordinating Committee (CTN-19-585462) grant to S.S.D. G.H. acknowledges the support of the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Training Program. The authors thank the following UCLA Broad Stem Cell core facilities: The Imaging core, Flow cytometry core, Genomics core, and the Stem Cell Banking core. The authors acknowledge the use of the Biological Nanostructures Laboratory within the California NanoSystems Institute for Illumina sequencing, supported by the University of California, Santa Barbara and the University of California, Office of the President . Human fetal tissue procurement in this research was supported by 5R24HD000836. Requests for fetal tissues can be made to Ian Glass at the University of Washington Birth Defects Laboratory.

Chapter 5 is unpublished work currently under peer review at Stem Cell Research. It reflects work performed by authors Grace V Hancock, Wanlu Liu, Lior Peretz, Di Chen, Joanna J Gell, Amanda J Collier, Jesse R Zamudio, Kathrin Plath, and Amander T Clark. LP assisted with the experiments. WL performed the data analysis. DC and JJG designed the foundational models and produced the critical preliminary data. JRZ assisted in the data analysis, and AJC in the naive

reversion experiments. KP designed the experiments and reviewed the manuscript, and ATC conceived and designed the experiments, maintained all institutional compliances, received the funding, and wrote the manuscript. The authors would like to thank Felicia Codrea, Jessica Scholes, and Jeffery Calimlim for FACS and Jinghua Tang for banking and culturing of the UCLA hESC lines. This work is supported by funds from an anonymous donor as well as NIH/NICHD R01 HD079546 (ATC), NIH/NIGMS P01 GM099134 (KP), and a Faculty Scholar grant from the Howard Hughes Medical Institute (KP). GH acknowledges the support of the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Training Program for supporting this work. Human fetal tissue research is supported by a grant to Ian Glass at the University of Washington Birth Defects laboratory 5R24HD000836-53. Human conceptus tissue requests can be made to bdrl@u.washington.edu.

VITA

Education

Johns Hopkins University
B.S. Cellular and Molecular Biology

Baltimore, MD
2015

University of California, Los Angeles
Ph.D. Molecular Biology

Los Angeles, CA
present

Publications

Hancock, G.V., Liu, W., Peretz, L., Chen, D., Gell, J.J., Collier, A.J., Zamudio, J.R., Plath, K., Clark, A.T. (exp. 2021) Divergent roles for KLF4 and TFCEP2L1 in Naive Ground State Pluripotency and Human Primordial Germ Cell Development. *In review*

Hancock, G.V., Wamaitha, S.E., Peretz, L., and Clark, A.T. (2021). Mammalian primordial germ cell specification. *Development* 148.

Gell, J.J., Liu, W., Sosa, E., Chialastri, A., **Hancock, G.**, Tao, Y., Wamaitha, S.E., Bower, G., Dey, S.S., and Clark, A.T. (2020). An Extended Culture System that Supports Human Primordial Germ Cell-like Cell Survival and Initiation of DNA Methylation Erasure. *Stem Cell Reports*.

Chen, D., Liu, W., Lukianchikov, A., **Hancock, G.V.**, Zimmerman, J., Lowe, M.G., Kim, R., Galic, Z., Irie, N., Surani, M.A., et al. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. *Biol. Reprod.* 97, 850–861.

Himburg, H.A., Doan, P.L., Quarmyne, M., Yan, X., Sasine, J., Zhao, L., **Hancock, G.V.**, Kan, J., Pohl, K.A., Tran, E., et al. (2017). Dickkopf-1 promotes hematopoietic regeneration via direct and niche-mediated mechanisms. *Nat Med* 23, 91–99.

EV-TRACK Consortium, Van Deun, J., Mestdagh, P., Agostinis, P., Akay, Ö., Anand, S., Anckaert, J., Martinez, Z.A., Baetens, T., Beghein, E., **et al.** (2017). EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research. *Nat Methods* 14, 228–232.

Zhao, Z., Muth, D.C., Mulka, K., Liao, Z., Powell, B.H., **Hancock, G.V.**, Metcalf Pate, K.A., and Witwer, K.W. (2020). miRNA profiling of primate cervicovaginal lavage and extracellular vesicles reveals miR-186-5p as a potential antiretroviral factor in macrophages. *FEBS Open Bio* 10, 2021–2039.

Selected Presentations

Hancock GV, Chen D, Peretz LP, Clark A. “The KLF family in human primordial germ cell development”. (2019, June 27). Invited talk at the International Society for Stem Cell Research Annual Meeting. Los Angeles, CA

Hancock GV, Chen D, Peretz LP, Clark A. “KLF4 and TFCEP2L1 in human primordial germ cell development”. (2018, September 23). Invited talk & Poster at From Stem Cells to Development Meeting, Surry, UK

Selected Honors & Awards

- 2017, 2018, 2019** Broad Stem Cell Research Center (BSCRC) Predoctoral fellowship
- 2016** National Science Foundation (NSF) Honorable Mention
- 2015** Graduation honors Johns Hopkins University
- 2014** Provost’s Undergraduate Research Award
- 2014** Second Decade Society summer internship grant
- 2013** Center for AIDS research summer scholar, top honors for best presentation

Leadership & Volunteer

Organizer Student Researchers United	Present
Senior New Ventures Fellow UCLA Technology Development Group	2019-2021
Senior Analyst & Mentor Tech Coast Angels	2020-2021
Financial Coordinator Advancing Women in Science and Engineering (AWiSE)	2016-2020
Communications Team Member Student Advisory Committee	2017-2018

Chapter 1- Introduction

Impact of the germline in human health

The germline is specified early in embryonic development and is comprised of the cells that will eventually give rise to the next generation. In animals, or metazoans, the first germline cells are called primordial germ cells (PGCs). These cells pass on both genetic identity and epigenetic information, and undergo a unique process known as reprogramming during their development. Simplified, reprogramming is when chromatin, DNA and its associated proteins, becomes more physically accessible in order to allow for widespread cell fate change. Given the crucial and complex nature of these cells, and the importance of the timing and processes by which they develop, it is not surprising that problems in human PGC (hPGC) specification, maintenance, and reprogramming can lead to infertility and developmental disorders. One in eight couples experience infertility, and about half are unresponsive to in vitro fertilization (Walker and Toblin, 2021). Additionally, germ cell tumors have an incidence of about 180,000 individuals in the US and are the most common adolescent tumor type in males (Fonseca et al., 2019; Lobo et al., 2019). Rare imprinting disorders such as Prader-Willi Syndrome, Angelman Syndrome, and others are caused by aberrant reprogramming and require both collective and individual attention to understand, treat, and prevent. An improved understanding of the processes of hPGC development would allow us to better inform couples and individuals of their fertility treatment options, avoiding expensive and demanding solutions that would be ineffective in certain cases. Additionally, understanding these mechanisms will contribute to development of diagnostic techniques and treatments to prevent, treat, and cure other diseases such as germ cell tumors and the rare disorders which lack treatments today.

hPGC specification

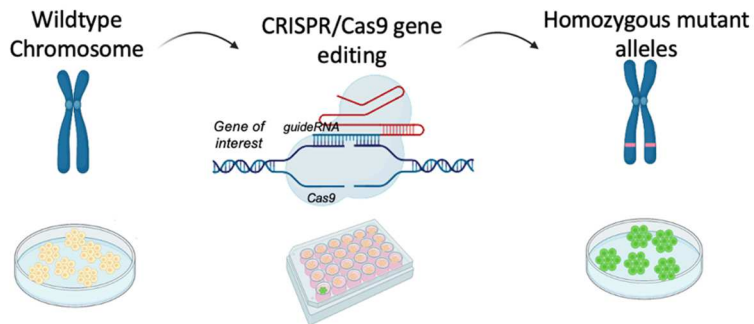
Generally, in more evolutionarily primitive species including model organisms *C. elegans*, *Drosophila*, and zebrafish, PGCs are specified by a process known as preformation (Extavour and Akam, 2003). Here, germ granules and plasma containing RNA and RNA-binding proteins are passed onto the PGCs while the rest of the somatic cells differentiate to give rise to the remainder of the organism. In mammalian PGC development, and interestingly other organisms such as the grasshopper (Chang, 2002), PGCs are specified by a process known as induction. In the induction model, PGCs are specified *de novo* by endogenous cues shortly after the specification of the epiblast, trophoblast (TE) and primitive endoderm (PE), the first three lineages to be specified in the developing embryo (Saitou and Yamaji, 2012; Sasaki et al., 2016). In Chapter 2, mammalian PGC development is reviewed in depth, comparing mouse, nonhuman primate, and human PGC specification. Inductive signaling, transcription factor control, epigenetic changes, and the state of the experimental models currently available to us in the field are also discussed.

In my work, to better understand hPGC specification, we relied on a disorganized aggregate model to induce hPGC-like cells (hPGCLCs) from human embryonic stem cells (hESCs). hESCs are first cultured for 24 hours in single-cell 2D-culture to induce differentiation towards incipient mesoderm-like cells (iMeLC), which are then dissociated and cultured as aggregates for 4 days (Sasaki et al., 2015). A percentage of cells within the aggregates successfully differentiates to hPGCLCs as the culture conditions include BMP4, a signaling molecule required for PGC induction conserved across many species (Hancock et al., 2021).

We used the hPGCLC *in vitro* system to identify genes and mechanistically study pathways specifically involved with hPGC development, as there remains notable species

divergence, especially between mouse and humans. For example, based on mouse models we know that the transcription factors TFAP2C, PRDM1 and PRDM14 are required for mouse PGC (mPGC) specification (Magnúsdóttir et al., 2013). To evaluate whether this is consistent in hPGCs, we used CRISPR/Cas9 gene editing to cause targeted deletion of genes of interest at the hESC stage to evaluate their role in specification as modeled by hPGCLC induction (Figure 1). Previously, SOX17 emerged as a novel transcription factor required for hPGC specification, while PRDM1 and PRDM14 were shown to be required and supportive, respectively (Gell et al., 2018; Irie et al., 2015; Sybirna et al., 2020). In Chapter 3, we further evaluated the role of transcription factors in hPGC specification and show that the transcription factors TFAP2C and EOMES are required for specification (Chen et al., 2017; Chen et al., 2018).

Figure 1. Strategy to develop mutant hESC lines. First, we transfected hESCs with plasmid DNA to target a gene of interest and encode Cas9. This is followed by a genotyping screen and expansion of colonies with the desired homozygous mutation.



Models of late-stage PGC development

The hPGCLC aggregate model is a robust system to represent newly specified, early hPGCs, albeit in a disorganized pattern. We can evaluate genetic and epigenetic identity as they are affected by various signaling molecules, growth factors, and genetic mutations. However, we do not see characteristics of later-stage PGCs, including expression of VASA or epigenetic changes. To study these later stages of maturation, and to generate substantial material for many

high-throughput methods, we require models where hPGCLCs can proliferate in culture and progress developmentally. In this work, we developed, optimized, and implemented a model where we can observe signs of hPGCLC maturation and proliferation including S-phase cell cycle activity as measured by Edu, and loss of H3K27me3 and UHRF1 (Gell et al., 2010). The development and characterization of this method is discussed in depth in chapter 4.

Pluripotency

During their transition from the early to the late stage, hPGCs re-acquire attributes that are characteristic of a state known as naive pluripotency (Chen et al., 2018). Generally speaking, pluripotency is the ability of a cell to give rise to multiple cell types, and is described in various states, defined *in vitro* as naive, formative, and primed, based on their relatedness to stages of the peri-implantation epiblast (Chan et al., 2013; Gafni et al., 2013; Hanna et al., 2010; Kinoshita et al., 2021; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014). The hESCs we use for induction to hPGCLCs are in the primed state and represent the post-implantation epiblast. The pre-implantation epiblast exists at a “ground state”, referring to its DNA, lacking repressive methylation and with permissive chromatin modifications, best represented by stem cells in the naive state. PGCs are unipotent, meaning they have one fate, but share notable similarities with the naive state including loss of DNA methylation, reactivation of the second X chromosome in females (Chitiashvili et al., 2020; Petropoulos et al., 2016; Vértésy et al., 2018), and upregulated expression of certain transcription factors including KLF4 and TFAP2L1 (Tang et al., 2015; Chen et al., 2018). Previous work showed a shared requirement for TFAP2C in the re-acquisition of naive pluripotency by primed hESCs (reversion), and also for hPGCLC induction (Chen et al., 2018; Pastor et al., 2018). Additionally, by investigating chromatin availability in hPGCs and

hPGCLCs using an Assay for Transposon-Accessible Chromatin sequencing (ATAC-seq), we observed highly significant differentially open binding motifs for the naive pluripotency factors of the KLF family and TFCP2L1. Given the role for TFAP2C in naive identity and upregulation of KLF4 and TFCP2L1 in hPGCs, we hypothesized that KLF4 and TFCP2L1 may have a role in hPGC development. In chapter 5 we investigated the transcription factors KLF4 and TFCP2L1 and consider how the pre-implantation epiblast modeled by naive hESCs is related to the molecular and functional identify of hPGCs.

References

- Chan, Y.-S., Göke, J., Ng, J.-H., Lu, X., Gonzales, K.A.U., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., and Ng, H.-H. (2013). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* *13*, 663–675.
- Chang, C., Dearden, P., and Akam, M. (2002). Germ line development in the grasshopper *Schistocerca gregaria*: vasa as a marker. *Dev Biol* *252*, 100–118.
- Chen, D., Liu, W., Lukianchikov, A., Hancock, G.V., Zimmerman, J., Lowe, M.G., Kim, R., Galic, Z., Irie, N., Surani, M.A., et al. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. *Biol. Reprod.* *97*, 850–861.
- Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E., et al. (2018). The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. *Cell Reports* *25*, 3591-3602.e5.
- Chitiashvili, T., Dror, I., Kim, R., Hsu, F.-M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K., et al. (2020). Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nature Cell Biology* *22*, 1436–1446.
- Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* *130*, 5869–5884.
- Fonseca, A., Frazier, A.L., and Shaikh, F. (2019). Germ Cell Tumors in Adolescents and Young Adults. *J Oncol Pract* *15*, 433–441.
- Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature* *504*, 282–286.
- Gell, J.J., Liu, W., Sosa, E., Chialastri, A., Hancock, G., Tao, Y., Wamaitha, S.E., Bower, G., Dey, S.S., and Clark, A.T. (2020). An Extended Culture System that Supports Human Primordial Germ Cell-like Cell Survival and Initiation of DNA Methylation Erasure. *Stem Cell Reports*.
- Hancock, G.V., Wamaitha, S.E., Peretz, L., and Clark, A.T. (2021). Mammalian primordial germ cell specification. *Development* *148*.
- Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. *Cell* *160*, 253–268.
- Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J., and Smith, A. (2021). Capture of Mouse and Human Stem Cells with Features of

Formative Pluripotency. *Cell Stem Cell* 28, 453-471.e8. Lobo, J., Gillis, A.J.M., Jerónimo, C., Henrique, R., and Looijenga, L.H.J. (2019). Human Germ Cell Tumors are Developmental Cancers: Impact of Epigenetics on Pathobiology and Clinic. *Int J Mol Sci* 20.

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

Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., et al. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat Cell Biol* 20, 553–564.

Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* 165, 1012–1026.

Saitou, M., and Yamaji, M. (2012). Primordial Germ Cells in Mice. *Cold Spring Harb Perspect Biol* 4, a008375.

Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., et al. (2015). Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* 17, 178–194.

Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Shiraki, N., Takakuwa, T., et al. (2016). The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Developmental Cell* 39, 169–185.

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

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

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* 15, 471–487.

Vértesy, Á., Arindrarto, W., Roost, M.S., Reinius, B., Torrens-Juaneda, V., Bialecka, M., Moustakas, I., Ariyurek, Y., Kuijk, E., Mei, H., et al. (2018). Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells. *Nature Communications* 9, 1873.

Walker, M.H., and Tobler, K.J. (2021). Female Infertility. In StatPearls, (Treasure Island (FL): StatPearls Publishing), p. Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., et al. (2014). Derivation of naive human embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* *111*, 4484–4489.

Chapter 2

Mammalian Primordial Germ Cell Specification

REVIEW

Mammalian primordial germ cell specification

Grace V. Hancock^{1,2,3}, Sissy E. Wamaitha^{1,3}, Lior Peretz¹ and Amander T. Clark^{1,2,3,4,*}**ABSTRACT**

The peri-implantation window of mammalian development is the crucial window for primordial germ cell (PGC) specification. Whereas pre-implantation dynamics are relatively conserved between species, the implantation window marks a stage of developmental divergence between key model organisms, and thus potential variance in the cell and molecular mechanisms for PGC specification. In humans, PGC specification is very difficult to study *in vivo*. To address this, the combined use of human and nonhuman primate embryos, and stem cell-based embryo models are essential for determining the origin of PGCs, as are comparative analyses to the equivalent stages of mouse development. Understanding the origin of PGCs in the peri-implantation embryo is crucial not only for accurate modeling of this essential process using stem cells, but also in determining the role of global epigenetic reprogramming upon which sex-specific differentiation into gametes relies.

KEY WORDS: Germ cells, Peri-implantation, Primordial germ cells, Embryo, Implantation, Pluripotency

Introduction

Germ cells are essential for inheritance because they are the only cell type capable of passing genetic and epigenetic information from parent to offspring. In some model organisms, such as *Drosophila*, *C. elegans*, *Xenopus laevis* and zebrafish, the germline is specified by a process known as pre-formation (Extavour and Akam, 2003). In these species, upon fertilization and cellularization/cleavage, the embryonic cells that inherit certain RNAs and proteins from the fertilized oocyte maintain germ cell fate, while the remaining cells differentiate into somatic cells. These first germline cells in the metazoan embryo are called primordial germ cells (PGCs). Unlike the model organisms that use pre-formation, mammals specify PGCs through a process called induction. One of the major distinguishing features of mammalian PGCs at the end of the PGC induction period is their unique epigenetic state characterized by global DNA demethylation, imprint erasure and, in female embryos, activation of both X chromosomes. Therefore, understanding the development of PGCs in mammals requires models that incorporate features that not only recapitulate the inductive signaling events, but also those that establish the characteristic epigenetic state.

In mammals, PGCs are induced after the differentiation of the trophectoderm (TE) and primitive endoderm (PE) (Saitou and Yamaji, 2012; Sasaki et al., 2016). This window of mammalian development is very difficult to study due to the small number of cells

in the embryo, combined with the dynamic changes in cell fate as the embryonic and extra-embryonic cell lineages heterogeneously develop to generate the conceptus. It is now appreciated that there are differences in the cell and molecular mechanisms involved in extra-embryonic lineage differentiation and PGC specification between mammals. For example, in mouse and pig (Box 1) models, PGCs are specified in a region of the peri-implantation embryo called the posterior proximal epiblast (Kobayashi et al., 2017; Lawson et al., 1999). However, in cynomolgus macaques (*Macaca fascicularis*; 'cyno'), PGCs are first specified in an extra-embryonic tissue called the amnion and possibly the posterior epiblast (Sasaki et al., 2016). Whether PGCs in other nonhuman primates and humans are specified similarly in the amnion or have a dual origin for specification is an ongoing debate in the field, yet to be clarified by emerging models and model systems.

Given these differences in PGC specification among mammals, it is useful to re-evaluate the transition of embryonic progenitor cells into PGCs in the available mammalian model organisms as well as stem cell-based embryo models. These comparisons are essential to identify the most appropriate model(s) for studying human PGC specification. In this Review, we discuss what is known about PGC specification in mouse, nonhuman primates and humans, drawing primarily from *in vivo* evidence. We highlight similarities and differences in PGC specification between these systems, as well as current gaps in our understanding, particularly relating to the unique process of epigenetic reprogramming in PGCs. Finally, we discuss how *in vitro* and *ex vivo* models can address major unanswered questions in PGC specification.

PGC specification in the mouse embryo


For the past three decades, the mouse has been used as the major mammalian model to study early embryo development, including mouse PGC (mPGC) specification. In the mouse, a tissue called the extra-embryonic ectoderm (ExE) is primarily responsible for providing the crucial signals that induce mPGC from the proximal epiblast (Lawson et al., 1999). Therefore, understanding the origins of the ExE, and its physical relationship to the mPGC precursors, is essential for interpreting data from new stem cell-based embryo models, as well as the specification of mPGCs *in vivo*. Although the ExE differentiates after implantation, the immediate precursors to the ExE are specified before the embryo implants (Tanaka et al., 2013; Rossant, 1988). In this section, we describe how the ExE arises during pre- and peri-implantation development in the mouse. We discuss the tissues and signals that contribute to specifying mPGC precursors in the posterior proximal epiblast, and finally the factors needed for mPGC specification from these precursors.

Development of the extra-embryonic ectoderm in mouse

The blastocyst is the last embryonic stage before embryo implantation. Immediately before implantation, at embryonic day (E) 4.5, the inner cell mass (ICM) differentiates into two distinct cell types: the pluripotent pre-implantation epiblast (pre-EPI) and a new epithelial layer called the primitive endoderm (PE) (Rossant et al.,

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

*Author for correspondence (clarka@ucla.edu)

 A.T.C., 0000-0003-2483-3982

Box 1. Pig primordial germ cell specification

The porcine (pig) embryo develops as a bilaminar disc during implantation similar to humans and monkeys (Keibel, 1897). Furthermore, porcine primordial germ cell (pPGC) precursors emerge from brachyury (T)-competent progenitors in the posterior post-epiblast (EPI) similar to mouse egg cylinder embryos (Kobayashi et al., 2017). Isolation of pig epiblasts reveals that post-EPI competency for pPGC specification requires Wnt (Kobayashi, et al., 2017). As detailed in this Review, BMP4 signaling from the extra-embryonic ectoderm (ExE) is required to specify mPGCs in egg cylinder embryos. In the pig embryo, BMP4 is expressed by extra-embryonic mesoderm (ExM) and nascent mesoderm cells (Magana et al., 2014). However, whether these tissues provide the source of BMP4 for pPGC specification has not yet been resolved. Once specified, pPGC precursors are identified as co-expressing the same transcription factors used to identify hPGCs. These include Sox17, Blimp1 and Nanog (Kobayashi et al., 2017). By E12.5, pPGC specification from precursors begins with expression of Tfp2c (Kobayashi et al., 2017). Competency of the post-EPI to specify additional pPGCs is reduced as gastrulation continues; however, given the lack of proliferation in newly specified pPGCs, it is speculated that recruitment of pPGCs from T-competent pPGC precursors continues until ~E15.5, 4 days after pPGC specification is initiated (Kobayashi et al., 2017).

2003; Chazaud et al., 2006). TE cells furthest from the embryonic pole become mural TE cells, which contact the uterine wall to trigger implantation (Dickson, 1963; Kirby et al., 1967; Armant, 2005; Sutherland, 2003; Hu and Cross, 2010). Conversely, the polar TE cells associated with the pre-EPI and PE remain multipotent and differentiate into ExE at around E5.5 (Copp, 1978).

As the ExE is specified, additional major morphological and cellular changes occur (Bardot and Hadjantonakis, 2020; Rossant et al., 2003). These changes include conversion of the pre-EPI to the early post-implantation EPI (post-EPI) and specialization of the PE epithelium into parietal endoderm, which lines the blastocoel cavity and visceral endoderm (VE) that encapsulates the ExE and post-EPI of the embryo (Rossant et al., 2003; Gardner and Rossant, 1979; Lawson and Pedersen, 1987). Cavitation within the early post-EPI occurs simultaneously with VE encapsulation, creating a polarized columnar epithelium in a cup-like shape containing a central pro-amniotic cavity. The post-EPI continues to expand, converting into a pseudostratified epithelium with the conceptus resembling a cylindrical structure (Coucovanis and Martin, 1995). At this time, the site of connection to the uterus defines the proximal pole of the proximodistal axis. The ExE also undergoes cavitation to create an ExE epithelium that is contiguous with the post-EPI and a cavity that becomes continuous with the amniotic cavity. The stages between E5.0, when the pro-amniotic cavity first appears, and E7.5, when the amnion and chorion are fully formed and gastrulation is almost complete, are called the 'egg cylinder stages' of mouse embryo development; it is during these stages that mPGCs are specified (Fig. 1A).

Specification of mouse PGCs

One of the earliest reports of mPGC specification identified alkaline phosphatase-positive mPGCs at E7.5, located outside the mouse embryo in the extra-embryonic mesoderm (ExM) (Ginsburg et al., 1990). However, using elegant transplantation experiments it was revealed that both the proximal and distal ends of the posterior post-EPI at E6.5 are equally competent for mPGC specification, provided that the epiblast is positioned adjacent to the ExE (Tam and Zhou, 1996), indicating that the specification of mPGC precursors is reliant on short-range signals from the adjacent ExE, before mPGCs migrate to the ExM (Saitou et al., 2002). Using lineage tracing, it has been

revealed that the mPGC precursors in the posterior post-EPI originate at E6.25 as a small number of cells positive for B-lymphocyte-induced maturation protein 1 (Blimp1; also known as Prdm1) adjacent to the ExE (Ohinata et al., 2005). Indeed, subsequent single-cell analyses have revealed that the Blimp1-positive mPGC precursors progress through a Sox2-negative Hoxb1-positive state, with essentially all of them remaining in the germline to re-express Sox2 and repress Hoxb1 (Kurimoto et al., 2008) (Fig. 1A). From around E7.25, mPGCs upregulate early onset PGC specification genes, including transcription factor AP-2 γ (Tfp2c), Prdm14 and Tdrd5 (Yamaji et al., 2008; Kurimoto et al., 2008; Weber et al., 2010). Upregulation of developmental pluripotency associated 3 (Dppa3; also known as Stella), tissue non-specific alkaline phosphatase (Tnap) and Nanos3 occur by E7.5 (Saitou et al., 2002; Kurimoto et al., 2008) (Table 1).

The major signaling molecule secreted by the ExE that acts on the posterior post-EPI to initiate mPGC specification is bone morphogenetic protein 4 (BMP4) (Lawson et al., 1999; Ohinata et al., 2009). During egg cylinder formation, BMP4 not only signals to the posterior post-EPI through canonical BMP receptors and the downstream Smad family of signal transducers (Zhao, 2003; Saitou and Yamaji, 2010; Senft et al., 2019), but also through activin A receptor type 1 (Acvr1; also known as Alk2) on VE cells to indirectly facilitate specification of the mPGC precursors (de Sousa Lopes et al., 2004). In addition, BMP8b (expressed by the ExE) and BMP2 (expressed by the VE) also regulate mPGC specification, as shown by a lack of or a severe reduction in the number of mPGCs in BMP8b- and BMP2-null mutant mice, respectively (Ying et al., 2000; Ying and Zhao, 2001).

Although mPGCs are specified in the posterior post-EPI, the entire epiblast could, in theory, be competent to develop into mPGCs if inhibitory signals are not produced by the anterior visceral endoderm (AVE) to restrict the field of BMP signaling to the posterior region (Ohinata et al., 2009). The AVE is a derivative of the PE, with BMP8b secretion from the ExE responsible for regulating AVE development (Ohinata et al., 2009). In addition, Wnt3 secreted by the posterior post-EPI is necessary to establish competency for mPGC specification, because Wnt3-deficient embryos fail to express Blimp1 or Prdm14 in the post-EPI, despite expression of BMP4 in the ExE (Ohinata et al., 2009). Wnt3 also activates the mesodermal-associated transcription factor brachyury (T) in post-EPI cells, with T regulating expression of Blimp1 (Aramaki et al., 2013). Further insights into the role of these signaling effectors have been established from *in vitro* model systems discussed in subsequent sections. For a discussion on the timing of PGC specification in the mouse relative to other species, see Box 2.

PGC specification in the monkey embryo

Models of peri-implantation and early post-implantation monkey embryo development have relied primarily on the use of rhesus and cyno macaques (*Macaca mulatta* and *Macaca fascicularis*) and, to a lesser extent, the common marmoset monkey (*Callithrix jacchus*). Morphological studies examining developmental stages from fertilization to formation of the blastocyst have suggested that the pre-implantation window is similar to the mouse, albeit with differences in timing (Nakamura et al., 2016; Boroviak and Nichols, 2017) (Box 2; Fig. 1B). Here, we describe and compare the stages of early development in rhesus, cyno and marmosets, highlighting similarities and differences from mouse development, as well as how these differences may impact the signaling environment in the peri-implantation embryo during primate PGC specification.

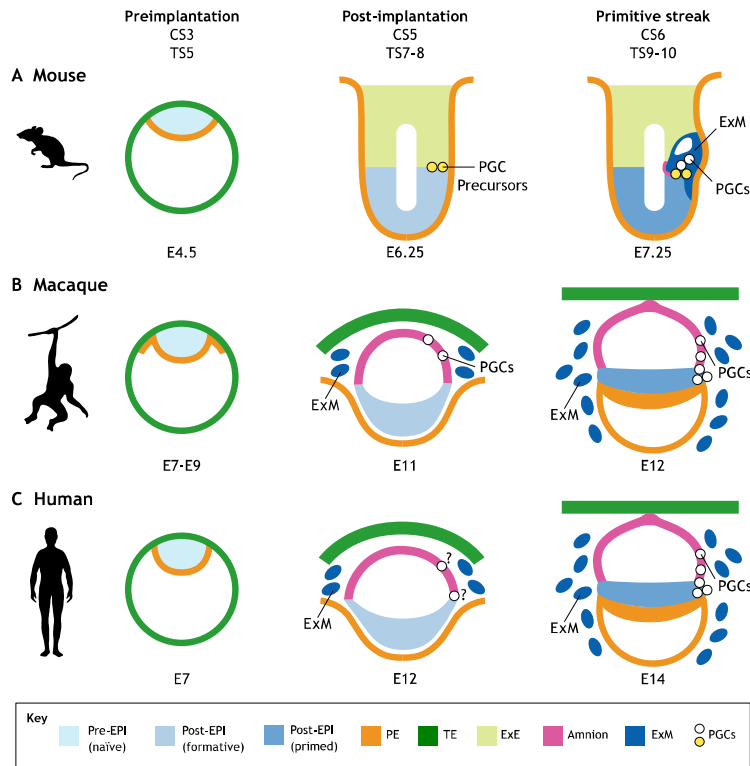


Fig. 1. Representation of peri-implantation embryo development in mammals. Mouse (A), monkey (B) and human (C) development. Formative pluripotency is described only in mouse and is therefore inferred in the other species. (A) In mouse embryos, the egg cylinder stages begin at Theiler stage (TS) 7. A pre-streak embryo is shown at embryonic day E6.25, which can be assigned to late TS8 with the specification of *Blimp1*⁺ primordial germ cell (PGC) precursors (yellow). Primitive streak formation begins at TS9, with mPGC specification from the precursors occurring late in TS9 and continuing in TS10. (B) In the monkey, cyPGCs are first identified at E11 [Carnegie stage (CS) 5c] in the amnion, with morphological evidence of primitive streak formation occurring in CS6. (C) In humans, *ex vivo* culture indicates that hPGCs can be identified before primitive streak formation at E12. '?' indicates that the cell layer of origin is not yet known for humans. At CS6, cyPGCs are found in the posterior primitive streak. In the mouse embryo, bone morphogenetic protein 4 (BMP4) and BMP8b are produced from extra-embryonic endoderm (ExE) to induce mPGC specification. In the nonhuman primate embryo, BMP4 is reported as being produced by amnion cells and possibly by extra-embryonic mesoderm (ExM). In the human embryo, the cell of origin is not yet known, and the source of BMP for inducing PGC specification is also unclear. Supporting embryonic and extra-embryonic tissue types are depicted as: preimplantation epiblast (pre-EPI), represented in culture by the naïve state; post-implantation epiblast (post-EPI), represented by the formative and primed states; primitive endoderm (PE); and trophoblast (TE).

Pre-implantation development in nonhuman primates

As in the mouse, the inner cell mass cells (ICMs) of nonhuman primate blastocysts segregate into pre-EPI and PE immediately before implantation (Heuser and Streeter, 1941; Enders and Schlafke, 1981; Nakamura et al., 2016). Studies in the rhesus have indicated that TE/PE/pre-EPI lineage segregation occurs between E5 and E7 (Enders and Schlafke, 1981; Liu et al., 2018) (Fig. 1B). In contrast, *in vitro* culture of cyno pre-implantation embryos suggests that this lineage segregation occurs between E7 and E9, indicating some differences in developmental timing between macaque species (Nakamura et al., 2016). Similar dynamics to cyno pre-implantation have also been observed in marmoset embryos, though further stages of early embryogenesis are then delayed compared with other primates (Boroviak et al., 2015, 2018; Phillips, 1976). Interestingly, in cyno embryos, the PE begins to expand to cover the blastocoel cavity prior to hatching from the zona pellucida, which is required for

implantation (Nakamura et al., 2016). This has not been observed in mice or humans (Fig. 1B).

Peri-implantation development in nonhuman primates

Despite relative similarities in the pre-implantation stages, the peri-implantation and early post-implantation stages are different between monkeys and mice. Starting at the point of embryo implantation into the uterine epithelium, it is the polar TE, rather than the mural TE, of monkey embryos that proliferates and differentiates into cytotrophoblast cells that then become the multinucleated invasive syncytiotrophoblast (Enders, 1989; Enders et al., 1983; Enders and Lopata, 1999; Smith et al., 1987; Wislocki and Bennett, 1943). Furthermore, as the embryo implants, monkey embryos develop into a bilaminar disc structure, rather than an egg cylinder (Boroviak and Nichols, 2017; Enders et al., 1986; Heuser and Streeter, 1941). Formation of the bilaminar disc involves cavitation of the post-EPI to

DEVELOPMENT

Table 1. Key features of primordial germ cell (PGC) development in different mammals

Species	Time: location of PGC precursor	Signals (source)	Factors needed for specification	Key marker genes expressed by PGCs
Mouse	E6.25; post-EPI	BMP4 and BMP8b (ExE); BMP2 (VE); and Wnt3 (post-EPI)	Prdm14, Prdm1, Tfap2c and Nanog	Precursor, <i>Prdm1</i> ; early, <i>Prdm14</i> , <i>Tfap2c</i> , <i>Tdrd5</i> , <i>Dppa3</i> , <i>Tnap</i> , <i>Nanog</i> and <i>Nanos3</i>
Pig	E11.5; post-EPI	BMP4 and Wnt (ExM and mesoderm)	Not tested	Precursor, <i>T</i> (brachyury); early, <i>SOX17</i> , <i>PRDM1</i> , <i>TFAP2C</i> , <i>NANOG</i> and <i>NANOS3</i>
Nonhuman primate	E11; dorsal amnion, and beneath EPI and VE	BMP4 (amnion*, ExM*), Wnt3a (amnion*, post-EPI*) and BMP2 (VE*)	Sox17 [†] , Tfap2c [†] and Prdm1 [†]	Precursor, <i>SOX17</i> and <i>TFAP2C</i> §; early, <i>PRDM1</i> and <i>T</i>
Human	E12; post-EPI, amnion [§]	BMP4 [‡] (unknown), activin A and Wnt [‡] (unknown)	Sox17 [†] , Prdm1 [†] , Tfap2c [†] and Prdm14 [‡]	Precursor, <i>TFAP2A</i> , <i>T</i> and <i>CDX2</i> ; early, <i>SOX17</i> , <i>TFAP2C</i> , <i>BLIMP1</i> , <i>NANOG</i> and <i>NANOS3</i>

*Proposed source as observed *in vivo* or *ex vivo*.

†Functional information from *in vitro* models.

§Proposed location of precursor.

ExE, extra-embryonic endoderm; ExM, extra-embryonic mesoderm; post-EPI, post-epiblast; VE, visceral endoderm.

form the amniotic cavity at around E10.5 in rhesus embryos, E11 in cyno embryos and by E16 in marmoset embryos (Enders et al., 1986; Moore, 1985). The post EPI-derived cells neighboring the polar TE become amnion cells, whereas the post-EPI cells on the floor of the amniotic cavity form the embryonic disc (Enders and King, 1988; Luckett, 1975; Moore, 1985). During amniotic cavity formation the PE cells continue to proliferate and line the mural TE with a layer of cells and extracellular matrix known as Heuser's membrane (Enders et al., 1986; Luckett, 1975; Sasaki et al., 2016). The PE cells will also form a secondary yolk sac structure that is unique to anthropoid primates (Luckett, 1978). Emergence of the ExM can be observed between E12 and E14 in the rhesus, and between E11 and E15 in cyno, immediately before and coincident with gastrulation (Enders and King, 1988; Luckett, 1978; Nakamura et al., 2016). In contrast to the mouse, where the ExM first emerges from the post-EPI, anatomical studies have suggested that the earliest monkey ExM originates from PE-derived visceral endoderm (Enders and King, 1988). However, this has yet to be confirmed using lineage-tracing methods. The rhesus and cyno ExM surrounds the bilaminar disc embryo in the space between the cytotrophoblast and the amniotic epithelial cells (Enders and King, 1988; Nakamura et al., 2016) (Fig. 1B).

Box 2. Primordial germ cell specification timing

The length of primordial germ cell (PGC) specification from PGC precursors differs within the mammalian clade and may be related to the differing lengths of gestation. In mPGC specification, the process of recruiting the *Blimp1*-positive mPGC precursors (first identified at E6.25) ceases around E7.5 (Ohinata et al., 2009). In the pig, pPGC precursors are first identified at E11.5, with recruitment estimated to end by E15.5 (Kobayashi et al., 2017). In humans and nonhuman primates, the timing of PGC specification is referenced in Carnegie stages (CS), a system by which embryo development is ordered based upon the appearance of morphological structures rather than time. Staging embryos using the CS system enables human development to be measured against other species, including monkeys, as well as the Theiler staging (TS) used for the mouse. Importantly, PGC specification is reported to begin in CS5 in humans and CS5c in monkeys. This timing loosely corresponds to human and monkey PGC specification, beginning around E11-E13. It remains to be determined when PGC specification ends in humans; however, given that cyPGCs are still detected in the amnion at E16 (Sasaki et al., 2016), it could be speculated that, similar to pPGC specification, recruitment from cyPGC precursors occurs over an extended time before the window of cyPGC specification finally closes.

Specification of nonhuman primate PGCs

Studying early post-implantation monkey embryos is scientifically challenging because of the highly specialized surgical and animal care requirements to work with pregnant monkeys in research. Yet careful single-cell immunofluorescence studies of early post-implantation cyno embryos have shown that cyno PGCs (cyPGCs) are first identified in the extra-embryonic dorsal amnion at E11. This discovery was based on co-expression of the cyPGC markers *Sox17* and *Tfap2c* in a subset of dorsal amnion cells (Sasaki et al., 2016). As described previously, in egg cylinder mouse embryos the ExE is essential to generating the source of BMP4 required for mPGC specification. However, a structure equivalent to the ExE is not present in any primate studied to date. Given the lack of ExE in monkey embryos, a major unanswered question is where is the location of the signaling source for primate PGC specification? Strong BMP4 expression has been detected throughout the amnion in a few embryos at E11, prior to the emergence of *Sox17/Tfap2c* cyPGCs, with expression shifting to the proximal/posterior amnion at E16. As determined in mice, *Wnt3* is necessary for cells to interpret the BMP4 signal and specify mPGCs (Ohinata et al., 2009); in monkey embryos, *Wnt3a* has been detected in the amnion at E11, as well as in the adjacent cytotrophoblast. Together, these results suggest that, in the absence of an ExE-like tissue, the amnion itself could be providing autocrine signals necessary for cyPGC specification. Alternatively, the ExM could be an extra-embryonic source of signals; weak BMP4 expression has been detected in the ExM of a single E11 embryo, although further investigations are required to draw robust conclusions (Sasaki et al., 2016). BMP2 has also been robustly detected in the VE at E11, which may indicate a putative role in cyPGC specification similar to mPGCs. Downstream targets of both BMP and Wnt signaling are expressed exclusively in the amnion compared to the epiblast at E11, while antagonists are expressed in the VE, suggesting that the amnion is the primary recipient of BMP signaling and further supporting the idea that the amnion is a source of cells at E11 that are receptive to signals for cyPGC specification (Sasaki et al., 2016) (Table 1).

In egg cylinder mouse embryos, mPGC specification is complete by E7.5 (Ohinata et al., 2009). In cyno embryos, cyPGCs can still be identified in the dorsal amnion at E17, 6 days after the first cyPGCs are identified, with additional cyPGCs located between the post-EPI and PE. Therefore, it could be speculated that cyPGC specification is an extended process similar to the pig (Box 2), with ongoing recruitment from PGC precursors before the window of PGC specification has

closed. Alternatively, PGCs detected at the later stages could be due to mis-migration, and this remains to be resolved. In addition to Sox17/Tfap2c double-positive cells, the competency factor *T* is also broadly expressed in the dorsal amnion around the time of cyPGC specification. However, it remains to be shown through lineage-tracing experiments whether the Sox17/Tfap2c double-positive cells in the amnion found prior to gastrulation correspond to the earliest cyPGCs. Furthermore, the relationship between these cells and those identified between the post-EPI and PE 24 hours (h) later remains to be determined.

PGC specification in the human embryo

Immediately prior to implantation, the ICM of human blastocysts also segregates into pre-EPI and PE lineages. Upon implantation, similar to nonhuman primates, human embryos develop into a bilaminar disc (Boroviak and Nichols, 2017; Cockburn and Rossant, 2010; Wamaitha and Niakan, 2018) (Fig. 1C). In order to study human development during implantation, the Carnegie and Kyoto collections of human embryos have been invaluable resources for the scientific community (Hertig et al., 1956; Hill, 2018; Nishimura et al., 1968; Müller and O'Rahilly, 1987; O'Rahilly and Müller, 2010; Yamaguchi and Yamada, 2018). In humans, embryo development is measured in Carnegie Stages (CS), which measures development based upon the ordered formation of key features, rather than a strict reliance on time after fertilization (Box 2). In human embryos, human PGCs (hPGCs) have been identified as early as CS5.

Human peri-implantation development

During CS5, human embryo implantation occurs and the amniotic sac containing the amnion and bilaminar embryonic disc are established. These events involve differentiation of the blastocyst TE cells into syncytiotrophoblast, which invades the maternal uterine wall, while the cytotrophoblast cells self-renew (Burton et al., 1999). Inside the implanting blastocyst, the pre-EPI and PE of the pre-implantation embryo transition into the post-EPI and PE of the bilaminar disc, with the amniotic sac and the primary yolk sac also forming (Hertig et al., 1956; Luckett, 1975). However, unlike the egg-cylinder mouse embryo, PE cells in the human embryo do not take on the role of surrounding the EPI and TE, and, similar to monkeys, there is no equivalent tissue to the ExE.

Specification of human PGCs

The first widely cited description of hPGCs in the post-implantation human embryo was in the yolk sac endoderm at ~E28, corresponding to CS12 (Witschi, 1948). However, additional histological evaluation of the 'Harvard 55' embryo in CS6a (~E13 post-fertilization) identified hPGCs 'stuffed with glycogen' near the edge of the embryonic disc (Hertig et al., 1958). There has been a single report of putative hPGCs at CS5 in the so-called 'Werner embryo' (~E12), where large round cells, presumed to be hPGCs, are identified in the post-EPI (Hertig et al., 1958; O'Rahilly and Müller, 2010). Single-cell transcriptome analyses of cells in the rostral, caudal and yolk sac regions of a single CS7 embryo has revealed that these earliest hPGCs are similar to PGCs in monkeys and pigs, and co-express the PGC marker genes *SOX17* and *NANOS3* (Tyser et al., 2020 preprint). The analysis of single embryos in each of these studies highlights the challenges of studying human development in the peri-implantation window. These analyses reinforce the reliance on comparative studies against other species and on new technologies in embryo culture, as well as stem cell models to study key developmental events that are specific to the germline and essential for human reproduction (Table 1; Fig. 1C).

DNA methylation remodeling and X-chromosome reactivation in PGCs

After specification, PGC development continues with epigenetic reprogramming. Completion of PGC epigenetic reprogramming occurs after entrance into the gonads and sex-specific differentiation. Here, we focus on two major epigenetic marks that are progressively removed soon after PGCs are induced: DNA methylation and the reactivation of the inactive X (Xi) in female embryos. Specifically in this section, we highlight the mouse embryo as a source of key information on the dynamics of DNA methylation remodeling. In addition, we highlight the need for stem cell and *ex vivo* embryo models to address key questions in DNA demethylation and X-chromosome reactivation, which cannot be easily studied using human tissue.

DNA methylation dynamics during PGC specification

The mouse has been used extensively to study the kinetics of DNA demethylation during and after mPGC specification (Guibert et al., 2012; Kobayashi et al., 2013; Seisenberger et al., 2012; Smith et al., 2012). In post-EPI mouse egg cylinder embryos ~70% of cytosines in a CpG sequence context are methylated (Seisenberger et al., 2012; Smith et al., 2012). DNA methylation levels have not been quantified at single-base resolution in committed mPGCs between E7.25 and E8.0. However, targeted sequencing of a small number of promoters at E8.5 suggests that mPGCs are highly methylated; by E9.5 CpG DNA methylation levels precipitously drop to ~30% (Guibert et al., 2012; Seisenberger et al., 2012; Seki et al., 2005). A further drop in DNA methylation is observed between E9.5 and E13.5, at which point the PGCs are located in the gonad (Kobayashi et al., 2013; Hill et al., 2018; Seisenberger et al., 2012), leading to the hypothesis that newly specified mPGCs inherit high levels of DNA methylation from the post-EPI, which is then erased from the mPGC genome during mPGC development.

In the monkey, immunofluorescence for 5-methylcytosine (5mC) has shown that, from at least E20 in the cyno (Sasaki et al., 2016) and from E28 in the rhesus (Sosa et al., 2018), the PGC genome has extremely low levels of 5mC relative to surrounding somatic cells. This is because 5mC has been oxidized to 5-hydroxymethyl cytosine (5hmC) (Sosa et al., 2018), a new epigenetic mark that serves as an intermediate in the DNA demethylation process. In the pig, pPGCs in the primitive streak have equivalent levels of 5mC to the soma, with 5mC demethylation occurring concomitantly with increasing levels of 5hmC at E14 (Kobayashi et al., 2017). In the early post-implantation human embryo, DNA demethylation has not been clearly defined and will require revisiting the international conversation on *ex vivo* culture of human embryos beyond 14 days post-fertilization and/or primitive streak formation (Hyun et al., 2016), which is when sufficient numbers of hPGCs could be recovered for analysis.

X-chromosome reactivation

In addition to DNA demethylation, female mPGCs also reactivate the Xi (de Sousa Lopes et al., 2008; Sangrithi et al., 2017; Sugimoto and Abe, 2007). X-chromosome inactivation (XCI) is associated with expression of *Xi-specific transcript* (*Xist*) (Brown et al., 1991). Around the time of mPGC specification, XCI is identified in >90% of post-EPI mouse cells (Rastan, 1982), and early reports indicated that *Xist* is expressed by mPGCs at E7.0 (Sugimoto and Abe, 2007). Reactivation of the Xi involves *Xist* repression in mPGCs after E9.5 (de Sousa Lopes et al., 2008; Sangrithi et al., 2017). In the pig, repression of *Xist* RNA and biallelic expression of X-linked genes occurs at E14 at the end of pPGC induction (Zhu et al., 2020). In

female hPGCs, there is now strong evidence that X-linked genes are biallelically expressed (Chitiashvili et al., 2020; Migeon and Jelalian, 1977; Vértessy et al., 2018), at a similar time to when promoter CGI methylation on the X chromosome is reported to be low (Tang et al., 2015). However, unlike female mPGCs, female hPGCs continue to express *XIST* RNA (Chitiashvili et al., 2020; Gkoutela et al., 2015).

Expression of *XIST* and biallelic expression of X-linked genes in hPGCs could represent a state of X-chromosome dose compensation, first described in human pre-implantation embryos, called X-chromosome dampening (XCD) (Chitiashvili et al., 2020; Petropoulos et al., 2016). XCD involves repression of X-linked genes from both X chromosomes and is associated with biallelic expression of *XIST*, as well as a hominid-specific long noncoding RNA called X active coating transcript (*XACT*) (Vallot et al., 2013, 2017). Recent work has revealed that *XIST* is expressed in female hPGCs whereas *XACT* is expressed in both male and female PGCs, with repression of *XIST* and *XACT* occurring as PGCs begin sex-specific differentiation (Chitiashvili et al., 2020). *XACT* is expressed in chimpanzees, but it is not found in more evolutionarily distant primates, such as macaques (Casanova et al., 2019). Therefore, studying X-chromosome dose compensation at the time of hPGC specification will rely on information from *in vitro* human pluripotent stem cell models, as well as *ex vivo* culture of human embryos beyond 14 days.

***In vitro* and *ex vivo* models of PGC specification**

Recent studies using pluripotent stem cell models of embryo development and *ex vivo* embryo culture have provided further insights into the cell and molecular requirements for PGC specification (Fig. 2A,B). In the mouse, culturing trophoblast stem cells (TSCs) with mouse embryonic stem cells (ESCs) results in the formation of an organized egg cylinder-like embryo structure called an ESC- and TSC-derived stem cell embryo (ETS-embryo), with mPGC specification occurring at the trophoblast/posterior-EPI boundary (Harrison et al., 2017) (Fig. 2B). These ETS-embryos do not contain VE, supporting the concept that the VE is not necessary for mPGC specification (Ohinata et al., 2009).

Mouse cell models of PGC development

ESCs are differentiated into PGC-like cells (PGCLCs) *in vitro*, a term that distinguishes them from endogenous PGCs. In the mouse, mPGCLC induction cannot occur directly from mouse ESCs cultured in 2i+LIF; instead, an intermediate epiblast-like cell (EpiLC) must first be induced (Hayashi et al., 2011). BMP4, BMP8 and other components are added to EpiLCs in order to robustly induce mPGCLCs within three-dimensional aggregates with an efficiency of ~13% (Hayashi et al., 2011, 2012) (Fig. 2A). Importantly, EpiLCs are considered equivalent to post-EPI cells in the E5.5 embryo, and are thought to represent a 'formative state' of pluripotency (Smith, 2017). In contrast, mouse ESCs cultured in 2i+LIF are equivalent to pre-EPI cells of the blastocyst, and this pluripotent state is referred to as the 'naïve' or 'ground state' of pluripotency. A third state of pluripotency known as 'primed' is represented by *in vitro* cultured cells derived from egg cylinder mouse embryos called epiblast stem cells (EpiSCs) (Brons et al., 2007; Kojima et al., 2014; Tesar et al., 2007; Wu et al., 2015). Mouse PGCLC induction from mEpiSCs occurs at significantly lower efficiency (1.5%) compared with EpiLCs (Hayashi et al., 2011; Hayashi and Surani, 2009; Huang et al., 2012). Therefore, the transition from naïve to formative pluripotency (but before acquiring primed pluripotency) is likely to be another essential requirement for mPGC competency.

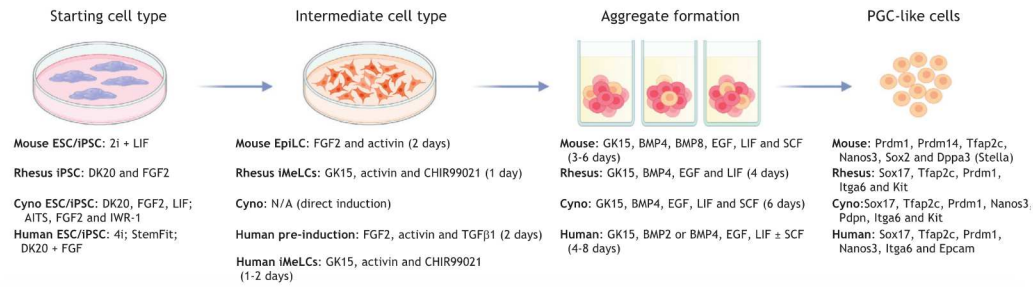
The differentiation of mPGCLCs from EpiLCs has revealed that ExE, ExM and VE cells are not required to specify mPGCLCs from male or female EpiLCs, provided that BMP4 and other molecules are present in the media (Hayashi et al., 2012, 2011). Using this *in vitro* system, the details of mPGC specification have been teased apart in even finer detail. For example, during mPGCLC induction, BMP4 and Wnt3 upregulate Blimp1 and T; the latter then directly targets and upregulates Blimp1, Prdm14 and Tfap2c for mPGC specification (Aramaki et al., 2013; Nakaki et al., 2013). Using this same model, overexpression of Prdm14, Blimp1 and Tfap2c in EpiLCs induces mPGCLC formation even if BMP4 is eliminated from the culture media (Aramaki et al., 2013; Magnúsdóttir et al., 2013). Doxycycline-induced overexpression of all three factors in the absence of cytokines causes robust expression of Blimp1 and Stella (34% of cells double positive) (Nakaki et al., 2013). Furthermore, induction of Prdm14 alone is capable of specifying mPGCLC fate in the absence of BMP signaling (3% Blimp1/Stella double-positive cells), as does Blimp1/Prdm14 (11.2%) and Blimp1/Tfap2c (2.6%) (Nakaki et al., 2013). Overexpression of the transcription factor Nanog can also induce mPGCLCs by promoting the expression of Blimp1 and Prdm14 (Murakami et al., 2016). With regard to DNA methylation, similar to the embryo, mPGCLCs are also highly methylated upon PGCLC specification and undergo X-chromosome reactivation (Hayashi et al., 2012; Shirane et al., 2016). In summary, combining information from *in vivo* and *in vitro* models of mPGC/PGCLC induction provides unprecedented details on the epigenome and on the relationships and networks of transcription factors that promote mPGC fate.

Nonhuman primate cell models of PGC development

In vitro nonhuman primate stem cell models

Monkey PGCLCs can also be induced from pluripotent stem cells using modified versions of the aggregate-based cytokine culture system developed in the mouse (Sakai et al., 2020; Sosa et al., 2018). Culture conditions include BMP4, but not BMP8, and use either an incipient-mesoderm intermediate (Sosa et al., 2018) or direct induction (Sakai et al., 2020). In the rhesus, cells triple positive for Sox17, Blimp1 and Tfap2c are observed in aggregates beginning 1 day after BMP4 exposure (Sosa et al., 2018). These three factors are crucial for both cyno and human PGC development (Irie et al., 2015; Sasaki et al., 2016). The newly specified rhesus PGCLCs have yet to initiate global epigenetic reprogramming and thus correspond to early *in vivo* PGCs, although partial reprogramming can be induced following xenotransplantation of rhesus PGCLCs into mouse testes (Sosa et al., 2018). In the cyno, Sox17, Blimp1 and Tfap2c are strongly upregulated in aggregates beginning 2 days after BMP4 exposure, with cyPGCLCs being transcriptionally similar to cyPGCs from E13 embryos onwards (Sakai et al., 2020) (Fig. 2A). Examining transcriptional dynamics during PGCLC induction in mice, cyno and human has revealed differences in gene expression between the mouse and the two primates, and further differences between human and cyno PGCLCs. For example, upregulation of Sox17, Blimp1 and Tfap2c occurs more rapidly during cyPGCLC induction compared with human, which likely explains the ability to dispense with the incipient-mesoderm intermediate, and genes associated with locomotion and migration are highly expressed in human PGCLCs compared with the cyno (Sakai et al., 2020). Although primate early embryogenesis models may show more similarity to humans, as would be predicted given the common bilaminar disc post-implantation embryo, these results indicate that there are still likely to be important differences.

A Stem cell-based aggregate models



B Stem cell-based embryo models

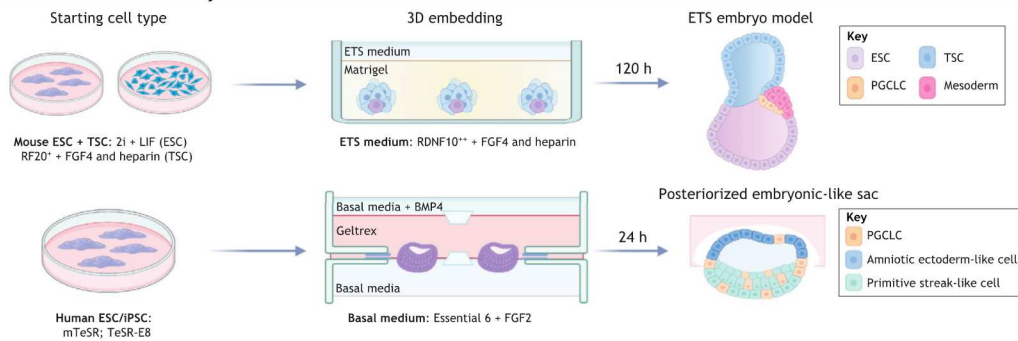


Fig. 2. Stem cell models of primordial germ cell induction. Modeling primordial germ cell (PGC) development with embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) uses either aggregate models or embryo models. (A) Aggregate models involve the generation of disorganized three-dimensional aggregates with primordial germ cell (PGC)-like cells (PGCLCs) induced from either a progenitor intermediate or by direct induction in response to BMP4 and other cytosines. Aggregate models have been created in the mouse (Hayashi et al., 2011, 2012), rhesus (Sosa et al., 2018), cyno (Sakai et al., 2020) and human (Irie et al., 2015; Sasaki et al., 2015; Chen et al., 2017, 2019). Growth factors, inhibitors and media types used for each PGCLC induction protocol are depicted. Proteins used to identify PGCLCs in each species are shown. (B) Embryo models involve modeling the spatial organization of some lineages in the peri-implantation embryo; these models have identified PGCLCs in the embryonic stem cell (ESC) and trophoblast stem cell (TSC) embryo model generated with mouse cells (Harrison et al., 2017), as well as in the posteriorized embryonic-like sac model generated with human cells (Zheng et al., 2019). AITS, 50% advanced RPMI, 50% neurobasal medium, and 1× insulin, transferrin and selenoprotein; BMP, bone morphogenetic protein; cyno, cynomolgus; DK20, DMEM/F-12 and 20% KSR; EGF, epidermal growth factor; EpiLC, epiblast-like cell; ETS, ES- and TS-derived stem cell embryo; FGF, fibroblast growth factor; GK15, GMEM and 15% KSR; iMeLCs, incipient mesoderm-like cell; iPSC, induced pluripotent stem cell; KSR, knockout serum replacer; mTeSR, commercial media; RDNF, 50% RPMI, 25% DMEM, 25% neurobasal medium and 10% fetal bovine serum (FBS); RF, RPM1 and 20% FBS; SCF, stem cell factor; TeSR-E8, TeSR-Essential 8 commercial media. Figure created with BioRender.com.

Ex vivo nonhuman primate embryo culture

Technological advances in *ex vivo* primate embryo attachment culture models are crucially needed for progress in understanding mechanisms common to monkey and human PGC development *ex vivo* and in particular DNA demethylation and reactivation of Xi. Monkey embryo attachment culture models have recently been developed based on techniques to generate peri-implantation stage mouse or human embryos from blastocysts (Bedzhov et al., 2014; Deglincerti et al., 2016; Shahbazi et al., 2016). Using this approach, two different studies have shown that cyno bilaminar disc embryos self-assemble entirely *in vitro* (Ma et al., 2019; Niu et al., 2019).

Ma and colleagues plated cyno blastocysts onto Matrigel-coated dishes and cultured them in normoxic conditions according to a modified method used for mouse embryo culture (Bedzhov et al., 2014; Ma et al., 2019). This results in embryo ‘attachment’ in two-thirds (67%) of cyno blastocysts by E9-E10, and bilaminar disc formation in about one-quarter (27%) of embryos by E13-E14.

Although the remaining embryos eventually attached by E11-E12, these had a much lower chance of successful disc formation. Meanwhile, Niu and colleagues used conditions that are more similar to the human embryo culture protocol (Deglincerti et al., 2016; Shahbazi et al., 2016). Cyno blastocysts were first cultured under hypoxia before plating onto tissue culture-treated plastic under normoxic conditions, following the rationale that, because *in vivo* embryos exist in a hypoxic environment, successful attachment may be better facilitated under these conditions (Niu et al., 2019). Indeed, here the majority (92%) of blastocysts attached by E10, forming a bilaminar disc by E13.

In both cases, cyno embryo attachment culture was performed up to E20, beyond the 14-day limit that applies to human embryos. Consequently, hallmarks of gastrulating cells could be observed (Niu et al., 2019). Ma and colleagues also observed neural structures in embryos at E19 (Ma et al., 2019). At earlier stages, Sox17/Tfap2c double-positive putative cyPGCs could be detected at E11-E12 in the

amnion, as well as beneath the EPI and in the VE (Ma et al., 2019). Indeed, this timing correlates with *in vivo* observations for cyPGC specification reported at CS5c (Sasaki et al., 2016). Niu and colleagues also observed Sox17/Tfap2c double-positive cyPGCs within the amnion at E13, and in the junction between the amnion and post-EPI at E14-E16 (Niu et al., 2019). Curiously, transcriptome analysis identified two clusters of cells with cyPGC-like identity: one that clustered with EPI cells and one that did not, indicating possible heterogeneity in cyPGC development within the *in vitro* attachment culture (Niu et al., 2019). Transcriptome analysis by Ma and colleagues also designated a population of putative early amnion cells expressing BMP4 (which might be providing PGC specification signals), while Niu and colleagues identified Wnt3 in the EPI at E13 and E17 (Ma et al., 2019; Niu et al., 2019); however, further study is required to identify possible signaling effectors in these models. Together, these two studies provide useful tools to interrogate the transcriptional profiles of primate peri-implantation lineages, and future studies could focus on understanding DNA demethylation in PGCs.

Human cell models of PGC development

Ex vivo human embryo culture

Similar to the cyno, *ex vivo* human embryo attachment culture has been crucial to understanding the cell and molecular events in peri-implantation and early post-implantation development in the absence of a uterus (Deglincerti et al., 2016; Shahbazi et al., 2016). Initial studies cultured human embryos to E13 *in vitro*, and noted morphological similarities to peri-implantation embryos at CS5, including the emergence of prospective amniotic and yolk sac cavities, although these could not be maintained (Deglincerti et al., 2016; Shahbazi et al., 2016). These first papers did not describe hPGCs, but subsequent work has used human *ex vivo* embryo attachment culture to investigate these cells (Chen et al., 2019; Popovic et al., 2019).

Chen and colleagues thawed 127 E5, E6 or E7 blastocyst-stage embryos and placed these into attachment culture under normoxic conditions. Eighty-five embryos (67%) successfully adhered and were cultured to E10, E11 or E12. Of these, 78 were imaged, and 26 embryos analyzed at E12. Cells co-expressing the PGC markers Nanog, Tfap2c and Sox17 were identified at E12 in 10% of embryos (2/26) where the three major lineages (post-EPI, VE and TE) could simultaneously be detected using immunofluorescence (Chen et al., 2019). Consistent with these observations, Popovic and colleagues have identified Oct4/Sox17/Ifitm3/Pdpn-positive, Sox2-negative cells at E12, designated as putative hPGCs. These cells were present in 34.8% of embryos (8/23), in which Oct4 (encoded by *POU5F1*) expression could be detected by immunofluorescence (Popovic et al., 2019). For this study, 141 E6 blastocyst stage embryos (cultured from pre-implantation embryos thawed at E2 or E3) were plated and cultured to E12. Forty-nine embryos (35%) were viable at E12, and 20 of these were excluded from further analysis because OCT4 expression could not be detected (Popovic et al., 2019). Further improvements to *ex vivo* human embryo attachment culture are required to not only increase the likelihood of successful attachment, but also to minimize embryo deterioration. Doing so requires a source of human embryos consented for research, which is not always practical for ongoing research activities given the ethical considerations surrounding human embryo research, coupled with funding restrictions in some parts of the world.

In vitro human stem cell models

Cell and molecular mechanisms involved in hPGC specification are studied using human ESCs (hESCs) and human induced pluripotent

stem cells (hiPSCs). Unlike mouse 2i+LIF ESCs, hESCs/hiPSCs are reported to be in the primed state of pluripotency given their similarity to mouse EpiSCs (Brons et al., 2007; Tesar et al., 2007). However, hESCs are derived from pre-implantation blastocysts, not post-implantation embryos, and their relatively flexible epigenome enables them to be reverted to the naïve ground state of pluripotency by simply changing the culture conditions, unlike mouse EpiSCs, which usually require transgene expression (Guo et al., 2009; Hanna et al., 2009; Silva et al., 2009; Takashima et al., 2014; Theunissen et al., 2014). Rare spontaneous reversions have been observed in a small proportion of dissociated single mEpiSCs cultured in stringent mouse ESC media; however, this varies widely between mEpiSC lines (Bao et al., 2009; Bernemann et al., 2011). Regardless, self-renewing mEpiSCs *in vitro* have low germline competency, while hESCs/hiPSCs can be differentiated efficiently into hPGCLCs (Irie et al., 2015; Hayashi et al., 2011; Sasaki et al., 2015). Collectively, this suggests that hESCs/hiPSCs and mEpiSCs are not functionally equivalent, even though they do share molecular characteristics. Instead, it is possible that hESCs/hiPSCs are more similar in function to mEpiLCs, which are highly competent for hPGCLC specification *in vitro*.

The first robust directed differentiation studies in which hESCs/hiPSCs were routinely differentiated into hPGCLCs revealed that, as in mouse, BMP4 is essential for specification. In addition, a new transcription factor network centered on *SOX17*, *PRDM1* and *TFAP2C* was found to be required for hPGCLC formation *in vitro* (Chen et al., 2017; Irie et al., 2015; Kojima et al., 2017; Sasaki et al., 2015). Prdm14 is also involved in hPGCLC induction; however, the downstream targets are different when compared with the mouse (Sybirna et al., 2020). These early studies involved differentiating hESCs/hiPSCs either directly in the presence of BMP4 and other cytokines, or after brief exposure to activin and Wnt agonists to induce an incipient mesoderm-like cell type, followed by exposure to BMP4 as disorganized three-dimensional (3D) aggregates. Single-cell RNA sequencing and pseudotime lineage tracing in this system has indicated that hPGCLCs are specified within 24–48 h of aggregate formation in the presence of BMP4 (Chen et al., 2019) (Fig. 2A). hPGCLC precursors emerge from a Wnt3/activin A-primed transitional pluripotent state, which has also been identified in the E12 attachment culture (Chen et al., 2019). In future studies, it will be interesting to determine whether the Wnt3/activin exposed pluripotent cells resemble the formative pluripotency of E5.5 mouse epiblasts.

After exposure to BMP4, pseudotime tracking predicts that the transcription factor Tfap2a becomes rapidly expressed in hPGC precursors, together with other lineage transcription factors traditionally associated with amnion, trophoblast and posterior primitive streak, including T and Cdx2 (Chen et al., 2019). Crucially, all these cell types differentiate in the presence of high levels of BMP4. Therefore, Tfap2a may simply serve as a read-out for high levels of BMP4 exposure; whether it is functionally involved in hPGCLC progenitor formation remains to be determined. hPGCLC specification next involves the rapid repression of Tfap2a, alongside transcription factors that mark the somatic lineage, such as Gata3 and Eomes. Once specified, hPGCLCs express Sox17, Tfap2c and Nanog, as in the embryo attachment system, as well as Blimp1 and Nanos3 (Chen et al., 2019). DNA methylation remodeling using the differentiation of naïve hESCs into EpiLCs, followed by hPGCLC differentiation has revealed similarities to the mouse (von Meyenn et al., 2016). However, studying XCR in hPGCLCs is challenged by the phenomenon of X-chromosome erosion (Mekhoubad et al., 2012), which is characterized by the abnormal and irreversible reactivation of

both X chromosomes (Mekhoubad et al., 2012; Patel et al., 2017). Improvement to naïve hESC/hiPSC stem cell culture could overcome this barrier in the future (Sahakyan et al., 2017).

More recently, stem cell-based embryo models have been developed. In these models, hESCs/hiPSCs are differentiated in microfluidics chambers to model amniotic sac formation (Zheng et al., 2019) or in 3D Matrigel to model luminogenesis of the amnion (Shahbazi et al., 2017, 2016). Gastruloids have also been modeled as 3D anteroposterior-organized aggregate structures, and bioengineered surfaces are used to model epiblast disc differentiation, demonstrating three germ layer specification and somitogenesis (Moris et al., 2020; Warmflash et al., 2014). The amniotic sac model (Fig. 2B) reveals that induction of hPGCLCs from hESCs occurs within 24 h of BMP4 exposure, with hPGCLCs first identified in presumptive amnion and pre-primitive streak-EPI. By 36 h, hPGCLCs are observed in much greater quantity in the EPI-like compartment. Given the amniotic sac model collapses between 36–72 h, new innovations are required to study the next steps in embryo development and hPGC formation.

Conclusions and future perspectives

Given the practicality of the mouse model, we have reached a comprehensive understanding of mPGC specification and epigenetic reprogramming. However, there are many apparent and important differences between rodent and primate PGC development. Through culturing human and primate embryos *ex vivo*, observing cyPGC, rhesus PGC and pig PGC development *in vivo*, and modeling human embryos *in vitro* using hESCs/hiPSCs, we are beginning to understand the cell(s) of origin for PGCs in primates and other bilaminar embryos relative to mice. Limitations to these methods include a scarcity of donated human tissue and high cost of nonhuman primate studies, resulting in low power and limited sample size in observational and high-throughput sequencing studies. Furthermore, *ex vivo* human embryo culture is limited by the ‘14-day’ rule and by embryo availability, and the crucial peri-implantation window cannot otherwise be observed in human development unless this rule is changed. Encouragingly, *in vitro* human pluripotent stem cell models are rapidly improving, although additional optimization is required to generate an integrated stem cell-based embryo model that represents all the cell types of the peri-implantation human embryo. That said, given the pace of this research area, it is expected that these techniques will be invaluable in understanding this developmental period.

Competing interests

The authors declare no competing or financial interests.

Funding

The authors' research was funded by grants from the National Institutes of Health/ Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD079546, R01HD098278 and R01HD058047). S.W. is supported by funds from the Iris Cantor-UCLA Women's Health Center Executive Advisory Board (NCATS UCLA CTS1 grant number UL1TR001881). Deposited in PMC for release after 12 months.

References

- Aramaki, S., Hayashi, K., Kurimoto, K., Ohta, H., Yabuta, Y., Iwanari, H., Mochizuki, Y., Hamakubo, T., Kato, Y., Shirahige, K. et al. (2013). A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev. Cell* **27**, 516–529. doi:10.1016/j.devcel.2013.11.001
- Armant, D. R. (2005). Blastocysts don't go it alone. Extrinsic signals fine-tune the intrinsic developmental program of trophoblast cells. *Dev. Biol.* **280**, 260–280. doi:10.1016/j.ydbio.2005.02.009
- Bao, S., Tang, F., Li, X., Hayashi, K., Gillich, A., Lao, K. and Surani, M. A. (2009). Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461**, 1292–1295. doi:10.1038/nature08534
- Bardot, E. S. and Hadjantonakis, A.-K. (2020). Mouse gastrulation: coordination of tissue patterning, specification and diversification of cell fate. *Mech. Dev.* **163**, 103617. doi:10.1016/j.mod.2020.103617
- Bedzhov, I., Leung, C. Y., Bialecka, M. and Zernicka-Goetz, M. (2014). In vitro culture of mouse blastocysts beyond the implantation stages. *Nat. Protoc.* **9**, 2732–2739. doi:10.1038/nprot.2014.186
- Bernemann, C., Greber, B., Ko, K., Sternecker, J., Han, D. W., Araújo-Bravo, M. J. and Schöler, H. R. (2011). Distinct developmental ground states of epiblast stem cell lines determine different pluripotency features. *Stem Cells* **29**, 1496–1503. doi:10.1002/stem.709
- Boroviak, T. and Nichols, J. (2017). Primate embryogenesis predicts the hallmarks of human naïve pluripotency. *Development* **144**, 175–186. doi:10.1242/dev.145177
- Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A. and Bertone, P. (2015). Lineage-specific profiling delineates the emergence and progression of naïve pluripotency in mammalian embryogenesis. *Dev. Cell* **35**, 366–382. doi:10.1016/j.devcel.2015.10.011
- Boroviak, T., Stirparo, G. G., Dietmann, S., Hernando-Herraez, I., Mohammed, H., Reik, W., Smith, A., Sasaki, E., Nichols, J. and Bertone, P. (2018). Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development* **145**, dev167833. doi:10.1242/dev.167833
- Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B., de Sousa Lopes, S. M. C., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195. doi:10.1038/nature05950
- Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R. and Willard, H. F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44. doi:10.1038/349038a0
- Burton, G. J., Jauniaux, E. and Watson, A. L. (1999). Maternal arterial connections to the placental intervillous space during the first trimester of human pregnancy: the Boyd collection revisited. *Am. J. Obstet. Gynecol.* **181**, 718–724. doi:10.1016/S0002-9378(99)70518-1
- Casanova, M., Moscatelli, M., Chauvière, L. É., Huret, C., Samson, J., Liyakat Ali, T. M., Rossopoff, O. and Rougeulle, C. (2019). A primate-specific retroviral enhancer wires the XACT lncRNA into the core pluripotency network in humans. *Nat. Commun.* **10**, 5652. doi:10.1038/s41467-019-13551-1
- Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell* **10**, 615–624. doi:10.1016/j.devcel.2006.02.020
- Chen, D., Liu, W., Lukianchikov, A., Hancock, G. V., Zimmerman, J., Lowe, M. G., Kim, R., Galic, Z., Irie, N., Surani, M. A. et al. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. *Biol. Reprod.* **97**, 850–861. doi:10.1093/biolre/iwx138
- Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W. et al. (2019). Human primordial germ cells are specified from lineage-primed progenitors. *Cell Rep.* **29**, 4568–4582.e5. doi:10.1016/j.celrep.2019.11.083
- Chitashvili, T., Dror, I., Kim, R., Hsu, F.-M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K. et al. (2020). Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nat. Cell Biol.* **22**, 1436–1446. doi:10.1038/s41556-020-00607-4
- Cockburn, K. and Rossant, J. (2010). Making the blastocyst: lessons from the mouse. *J. Clin. Invest.* **120**, 995–1003. doi:10.1172/JCI41229
- Copp, A. J. (1978). Interaction between inner cell mass and trophectoderm of the mouse blastocyst. I. A study of cellular proliferation. *J. Embryol. Exp. Morphol.* **48**, 109–125.
- Coucounavis, E. and Martin, G. R. (1995). Signals for death and survival: a two-step mechanism for cavitation in the vertebrate embryo. *Cell* **83**, 279–287. doi:10.1016/0092-8674(95)90169-8
- Deglicenti, A., Croft, G. F., Pietila, L. N., Zernicka-Goetz, M., Siggia, E. D. and Brivanlou, A. H. (2016). Self-organization of the in vitro attached human embryo. *Nature* **533**, 251–254. doi:10.1038/nature17948
- de Sousa Lopes, S. M. C., Roelen, B. A. J., Monteiro, R. M., Emmens, R., Lin, H. Y., Li, E., Lawson, K. A. and Mummery, C. L. (2004). BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **18**, 1838–1849. doi:10.1101/gad.294004
- de Sousa Lopes, S. M. C., Hayashi, K., Shovlin, T. C., Mifsud, W., Surani, M. A. and McLaren, A. (2008). X chromosome activity in mouse XX primordial germ cells. *PLoS Genet.* **4**, e30. doi:10.1371/journal.pgen.0040030
- Dickson, A. D. (1963). TROPHOBLASTIC GIANT CELL TRANSFORMATION OF MOUSE BLASTOCYSTS. *J. Reprod. Fertil.* **6**, 465–466. doi:10.1530/jrf.0.0060465
- Enders, A. C. (1989). Trophoblast differentiation during the transition from trophoblastic plate to lacunar stage of implantation in the rhesus monkey and human. *Am. J. Anat.* **186**, 85–98. doi:10.1002/aja.1001860107

- Enders, A. C. and King, B. F. (1988). Formation and differentiation of extraembryonic mesoderm in the rhesus monkey. *Am. J. Anat.* **181**, 327-340. doi:10.1002/aja.1001810402
- Enders, A. C. and Lopata, A. (1999). Implantation in the marmoset monkey: expansion of the early implantation site. *Anat. Rec.* **256**, 279-299. doi:10.1002/(SICI)1097-0185(19991101)256:3<279::AID-AR7>3.0.CO;2-O
- Enders, A. C. and Schlafke, S. (1981). Differentiation of the blastocyst of the rhesus monkey. *Am. J. Anat.* **162**, 1-21. doi:10.1002/aja.1001620102
- Enders, A. C., Hendrickx, A. G. and Schlafke, S. (1983). Implantation in the rhesus monkey: initial penetration of endometrium. *Am. J. Anat.* **167**, 275-298. doi:10.1002/aja.1001670302
- Enders, A. C., Schlafke, S. and Hendrickx, A. G. (1986). Differentiation of the embryonic disc, amnion, and yolk sac in the rhesus monkey. *Am. J. Anat.* **177**, 161-185. doi:10.1002/aja.1001770205
- Extavour, C. G. and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* **130**, 5869-5884. doi:10.1242/dev.00804
- Gardner, R. L. and Rossant, J. (1979). Investigation of the fate of 4*5 day post-coitum mouse inner cell mass cells by blastocyst injection. *J. Embryol. exp. Morph.* **52**, 141-152.
- Ginsburg, M., Snow, M. H. and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521-528.
- Gkountela, S., Zhang, K. X., Shafiq, T. A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y. and Clark, A. T. (2015). DNA demethylation dynamics in the human prenatal germline. *Cell* **161**, 1425-1436. doi:10.1016/j.cell.2015.05.012
- Guibert, S., Forné, T. and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.* **22**, 633-641. doi:10.1101/gr.130997.111
- Guo, G., Yang, J., Nichols, J., Hall, J. S., Eyres, I., Mansfield, W. and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063-1069. doi:10.1242/dev.030957
- Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A. W., Cassidy, J. P., Staerk, J., Carey, B. W., Lengner, C. J., Foreman, R., Love, J. et al. (2009). Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**, 513-524. doi:10.1016/j.stem.2009.04.015
- Harrison, S. E., Sozen, B., Christodoulou, N., Kyprianou, C. and Zernicka-Goetz, M. (2017). Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis *in vitro*. *Science* **356**, eaal1810. doi:10.1126/science.aal1810
- Hayashi, K. and Surani, M. A. (2009). 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
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532. doi:10.1016/j.cell.2011.06.052
- Hayashi, K., Ogushi, S., Kurimoto, K., Shimamoto, S., Ohta, H. and Saitou, M. (2012). Offspring from oocytes derived from *in vitro* primordial germ cell-like cells in mice. *Science* **338**, 971-975. doi:10.1126/science.1226889
- Hertig, A. T., Rock, J. and Adams, E. C. (1956). A description of 34 human ova within the first 17 days of development. *Am. J. Anat.* **98**, 435-493. doi:10.1002/aja.1000980306
- Hertig, A. T., Adams, E. C., McKay, D. G., Rock, J., Mulligan, W. J. and Menkin, M. F. (1958). A thirteen-day human ovum studied histochemically. *Am. J. Obstet. Gynecol.* **76**, 1025-1038; discussion 1040-1043. doi:10.1016/0002-9378(58)90185-6
- Heuser, C. and Streeter, G. (1941). Development of the macaque embryo. *Contributions to Embryology* **29**, 15-55.
- Hill, M. A. (2018). Two web resources linking major human embryology collections worldwide. *Cells Tissues Organs* **205**, 293-302. doi:10.1159/000495619
- Hill, P. W. S., Leitch, H. G., Requena, C. E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S. et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* **555**, 392-396. doi:10.1038/nature25964
- Hu, D. and Cross, J. C. (2010). Development and function of trophoblast giant cells in the rodent placenta. *Int. J. Dev. Biol.* **54**, 341-354. doi:10.1387/ijdb.082768dh
- Huang, Y., Osorno, R., Tsakiridis, A. and Wilson, V. (2012). *In vivo* differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Rep.* **2**, 1571-1578. doi:10.1016/j.celrep.2012.10.022
- Hyun, I., Wilkerson, A. and Johnston, J. (2016). Embryology policy: revisit the 14-day rule. *Nature* **533**, 169. doi:10.1038/533169a
- Irie, N., Weinberger, L., Tang, W. W. C., Kobayashi, T., Viukov, S., Manor, Y. S., Dietmann, S., Hanna, J. H. and Surani, M. A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253-268. doi:10.1016/j.cell.2014.12.013
- Keibel, F. (1897). *Normentafel zur Entwicklungsgeschichte des Schweines, Sus scrofa domestica*. Fischer.
- Kirby, D. R. S., Potts, D. M. and Wilson, I. B. (1967). On the orientation of the implanting blastocyst. *Development* **17**, 527-532.
- Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T. et al. (2013). High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616-627. doi:10.1101/gr.148023.112
- Kobayashi, T., Zhang, H., Tang, W. W. C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S., Contreras, D. A., Webb, R. et al. (2017). Principles of early human development and germ cell program from conserved model systems. *Nature* **546**, 416-420. doi:10.1038/nature22812
- Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A. F., Jones, V., Hor, A., de Alencastro, G., Logan, G. J. et al. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107-120. doi:10.1016/j.stem.2013.09.014
- Kojima, Y., Sasaki, K., Yokobayashi, S., Sakai, Y., Nakamura, T., Yabuta, Y., Nakaki, F., Nagaoka, S., Woltjen, K., Hotta, A. et al. (2017). Evolutionarily distinctive transcriptional and signaling programs drive human germ cell lineage specification from pluripotent stem cells. *Cell Stem Cell* **21**, 517-532.e5. doi:10.1016/j.stem.2017.09.005
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K. and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22**, 1617-1635. doi:10.1101/gad.1649908
- Lawson, K. A. and Pedersen, R. A. (1987). Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse. *Development* **101**, 627-652.
- Lawson, K. A., Dunn, N. R., Roelen, B. A. J., Zeinstra, L. M., Davis, A. M., Wright, C. V. E., Korving, J. P. W. F. M. and Hogan, B. L. M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424-436. doi:10.1101/gad.13.4.424
- Liu, D., Wang, X., He, D., Sun, C., He, X., Yan, L., Li, Y., Han, J.-D. J. and Zheng, P. (2018). Single-cell RNA-sequencing reveals the existence of naive and primed pluripotency in pre-implantation rhesus monkey embryos. *Genome Res.* **28**, 1481-1493. doi:10.1101/gr.233437.117
- Luckett, W. P. (1975). The development of primordial and definitive amniotic cavities in early rhesus monkey and human embryos. *Am. J. Anat.* **144**, 149-167. doi:10.1002/aja.1001440204
- Luckett, W. P. (1978). Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos. *Am. J. Anat.* **152**, 59-97. doi:10.1002/aja.1001520106
- Ma, H., Zhai, J., Wan, H., Jiang, X., Wang, X., Wang, L., Xiang, Y., He, X., Zhao, Z.-A., Zhao, B. et al. (2019). *In vitro* culture of cynomolgus monkey embryos beyond early gastrulation. *Science* **366**, eaax7890. doi:10.1126/science.aax7890
- Magaña, V. G., Rodríguez, A., Zhang, H., Webb, R. and Alberio, R. (2014). Paracrine effects of embryo-derived FGF4 and BMP4 during pig trophoblast elongation. *Dev. Biol.* **387**, 15-27.
- Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B. and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* **15**, 905-915. doi:10.1038/ncb2798
- Mekhoubad, S., Bock, C., de Boer, A. S., Kiskinis, E., Meissner, A. and Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* **10**, 595-609. doi:10.1016/j.stem.2012.02.014
- Migeon, B. R. and Jalilian, K. (1977). Evidence for two active X chromosomes in germ cells of female before meiotic entry. *Nature* **269**, 242-243. doi:10.1038/269242a0
- Moore, N. W. (1985). The use of embryo transfer and steroid hormone replacement therapy in the study of prenatal mortality. *Theriogenology* **23**, 121-128. doi:10.1016/0093-691X(85)90077-9
- Moris, N., Anlas, K., van den Brink, S. C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A. and Martínez Arias, A. (2020). An *in vitro* model of early anteroposterior organization during human development. *Nature* **582**, 410-415. doi:10.1038/s41586-020-2383-9
- Müller, F. and O'Rahilly, R. (1987). The development of the human brain, the closure of the caudal neuropore, and the beginning of secondary neurulation at stage 12. *Anat. Embryol.* **176**, 413-430. doi:10.1007/BF00310083
- Murakami, K., Günesdogan, U., Zyllicz, J. J., Tang, W. W. C., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S. and Azim Surani, M. (2016). NANOG alone induces germ cells in primed epiblast *in vitro* by activation of enhancers. *Nature* **529**, 403-407. doi:10.1038/nature16480
- Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y. and Saitou, M. (2013). Induction of mouse germ-cell fate by transcription factors *in vitro*. *Nature* **501**, 222-226. doi:10.1038/nature12417
- Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Yamamoto, T. and Saitou, M. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57-62. doi:10.1038/nature19096
- Nishimura, H., Takano, K., Tanimura, T. and Yasuda, M. (1968). Normal and abnormal development of human embryos: first report of the analysis of 1,213 intact embryos. *Teratology* **1**, 281-290. doi:10.1002/tera.1420010306
- Niu, Y., Sun, N., Li, C., Lei, Y., Huang, Z., Wu, J., Si, C., Dai, X., Liu, C., Wei, J. et al. (2019). Dissecting primate early post-implantation development using long-term *in vitro* embryo culture. *Science* **366**, eaaw5754. doi:10.1126/science.aaw5754

- Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S. C., Obukhanych, T., Nussenzweig, M., Tarakhovskiy, A. et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213. doi:10.1038/nature03813
- Ohinata, Y., Ohta, H., Shigetani, M., Yamanaka, K., Wakayama, T. and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571-584. doi:10.1016/j.cell.2009.03.014
- O'Rahilly, R. and Müller, F. (2010). Developmental stages in human embryos: revised and new measurements. *Cells Tissues Organs* **192**, 73-84. doi:10.1159/000289817
- Patel, S., Bonora, G., Sahakyan, A., Kim, R., Chronis, C., Langerman, J., Fitz-Gibbon, S., Rubbi, L., Skelton, R. J. P., Ardehali, R. et al. (2017). Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep.* **18**, 54-67. doi:10.1016/j.celrep.2016.11.054
- Petropoulos, S., Edsgård, D., Reinus, B., Deng, Q., Panula, S. P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R. and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012-1026. doi:10.1016/j.cell.2016.03.023
- Phillips, I. R. (1976). The embryology of the common marmoset (*Callithrix jacchus*). *Adv. Anat. Embryol. Cell Biol.* **52**, 3-47.
- Popovic, M., Bialecka, M., Gomes Fernandes, M., Taelman, J., Van Der Jeught, M., De Sutter, P., Heindryckx, B. and De Sousa Lopes, S. M. C. (2019). Human blastocyst outgrowths recapitulate primordial germ cell specification events. *Mol. Hum. Reprod.* **25**, 519-526. doi:10.1093/molehr/gaz035
- Rastan, S. (1982). Timing of X-chromosome inactivation in postimplantation mouse embryos. *J. Embryol. Exp. Morphol.* **71**, 11-24.
- Rossant, J. (1988). Development of extraembryonic cell lineages in the mouse embryo. In *Experimental Approaches to Mammalian Embryonic Development* (ed. J. Rossant, and R. A. Pedersen) pp. 97-120. London: Cambridge University Press.
- Rossant, J., Chazaud, C. and Yamanaka, Y. (2003). Lineage allocation and asymmetries in the early mouse embryo. *Philos. Trans. R. Soc. B Biol. Sci.* **358**, 1341-1349. doi:10.1098/rstb.2003.1329
- Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T. W., Kuoy, E., Langerman, J., Clark, A. T., Jaenisch, R. et al. (2017). Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87-101. doi:10.1016/j.stem.2016.10.006
- Saitou, M. and Yamaji, M. (2010). Germ cell specification in mice: signaling, transcription regulation, and epigenetic consequences. *Reproduction* **139**, 931-942. doi:10.1530/REP-10-0043
- Saitou, M. and Yamaji, M. (2012). Primordial germ cells in mice. *Cold Spring Harb. Perspect. Biol.* **4**, a008375. doi:10.1101/cshperspect.a008375
- Saitou, M., Barton, S. C. and Surani, M. A. (2002). A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300. doi:10.1038/nature00927
- Sakai, Y., Nakamura, T., Okamoto, I., Gyobu-Motani, S., Ohta, H., Yabuta, Y., Tsukiyama, T., Iwatani, C., Tsuchiya, H., Ema, M. et al. (2020). Induction of the germ cell fate from pluripotent stem cells in cynomolgus monkeys. *Biol. Reprod.* **102**, 620-638. doi:10.1093/biolre/foz205
- Sangrithi, M. N., Royo, H., Mahadevaiah, S. K., Ojarikre, O., Bhaw, L., Sesay, A., Peters, A. H. F. M., Stadler, M. and Turner, J. M. A. (2017). Non-canonical and sexually dimorphic X dosage compensation states in the mouse and human germline. *Dev. Cell* **40**, 289-301.e3. doi:10.1016/j.devcel.2016.12.023
- Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H. et al. (2015). Robust *in vitro* induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* **17**, 178-194. doi:10.1016/j.stem.2015.06.014
- Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Shiraki, N., Takakuwa, T. et al. (2016). The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Dev. Cell* **39**, 169-185. doi:10.1016/j.devcel.2016.09.007
- Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W. and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849-862. doi:10.1016/j.molcel.2012.11.001
- Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M. and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278**, 440-458. doi:10.1016/j.ydbio.2004.11.025
- Senft, A. D., Bikoff, E. K., Robertson, E. J. and Costello, I. (2019). Genetic dissection of Nodal and Bmp signalling requirements during primordial germ cell development in mouse. *Nat. Commun.* **10**, 1089. doi:10.1038/s41467-019-09052-w
- Shahbazi, M. N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., Fogarty, N. M. E., Campbell, A., Devito, L. G., Ilic, D. et al. (2016). Self-organization of the human embryo in the absence of maternal tissues. *Nat. Cell Biol.* **18**, 700-708. doi:10.1038/ncb3347
- Shahbazi, M. N., Scialdone, A., Skorupska, N., Weberling, A., Recher, G., Zhu, M., Jedrusik, A., Devito, L. G., Noli, L., Macaulay, I. C. et al. (2017). Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* **552**, 239-243. doi:10.1038/nature24675
- Shirane, K., Kurimoto, K., Yabuta, Y., Yamaji, M., Satoh, J., Ito, S., Watanabe, A., Hayashi, K., Saitou, M. and Sasaki, H. (2016). Global landscape and regulatory principles of DNA methylation reprogramming for germ cell specification by mouse pluripotent stem cells. *Dev. Cell* **39**, 87-103. doi:10.1016/j.devcel.2016.08.008
- Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737. doi:10.1016/j.cell.2009.07.039
- Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365-373. doi:10.1242/dev.142679
- Smith, C. A., Moore, H. D. M. and Hearn, J. P. (1987). The ultrastructure of early implantation in the marmoset monkey (*Callithrix jacchus*). *Anat. Embryol.* **175**, 399-410. doi:10.1007/BF00309853
- Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A. and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-344. doi:10.1038/nature10960
- Sosa, E., Chen, D., Rojas, E. J., Hennebold, J. D., Peters, K. A., Wu, Z., Lam, T. N., Mitchell, J. M., Sukhwani, M., Taylor, R. C. et al. (2018). Differentiation of primate primordial germ cell-like cells following transplantation into the adult gonadal niche. *Nat. Commun.* **9**, 5339. doi:10.1038/s41467-018-07740-7
- Sugimoto, M. and Abe, K. (2007). X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLoS Genet.* **3**, e116. doi:10.1371/journal.pgen.0030116
- Sutherland, A. (2003). Mechanisms of implantation in the mouse: differentiation and functional importance of trophoblast giant cell behavior. *Dev. Biol.* **258**, 241-251. doi:10.1016/S0012-1606(03)00130-1
- Sybirna, A., Tang, W. W. C., Pierson Smela, M., Dietmann, S., Gruhn, W. H., Brosh, R. and Surani, M. A. (2020). A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. *Nat. Commun.* **11**, 1282. doi:10.1038/s41467-020-15042-0
- Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W. et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254-1269. doi:10.1016/j.cell.2014.08.029
- Tam, P. P. L. and Zhou, S. X. (1996). The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev. Biol.* **178**, 124-132. doi:10.1006/dbio.1996.0203
- Tanaka, S. S., Nakane, A., Yamaguchi, Y. L., Terabayashi, T., Abe, T., Nakao, K., Asashima, M., Steiner, K. A., Tam, P. P. L. and Nishinakamura, R. (2013). Dullard/Ctdn1 modulates WNT signaling activity for the formation of primordial germ cells in the mouse embryo. *PLoS ONE* **8**, e57428. doi:10.1371/journal.pone.0057428
- Tang, W. W. C., Dietmann, S., Irie, N., Leitch, H. G., Floros, V. I., Bradshaw, C. R., Hackett, J. A., Chinnery, P. F. and Surani, M. A. (2015). A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161**, 1453-1467. doi:10.1016/j.cell.2015.04.053
- Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 169-199. doi:10.1038/nature05972
- Theunissen, T. W., Powell, B. E., Wang, H., Mitalipova, M., Faddah, D. A., Reddy, J., Fan, Z. P., Maetzel, D., Ganz, K., Shi, L. et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471-487. doi:10.1016/j.stem.2014.07.002
- Tyzer, R. C. V., Mohammadov, E., Nakanoh, S., Vallier, L., Scialdone, A. and Srinivas, S. (2020). A spatially resolved single cell atlas of human gastrulation. *bioRxiv* 2020.07.21.213512.
- Vallot, C., Huret, C., Lesecque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A., Duret, L. and Rougeulle, C. (2013). XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45**, 239-241. doi:10.1038/ng.2530
- Vallot, C., Patrat, C., Collier, A. J., Huret, C., Casanova, M., Liyakat Ali, T. M., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P. J. et al. (2017). XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell Stem Cell* **20**, 102-111. doi:10.1016/j.stem.2016.10.014
- Vértesy, Á., Arindrarto, W., Roost, M. S., Reinus, B., Torrens-Juaneda, V., Bialecka, M., Moustakas, I., Ariyurek, Y., Kuijk, E., Mei, H. et al. (2018). Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells. *Nat. Commun.* **9**, 1873. doi:10.1038/s41467-018-04215-7
- von Meyenn, F., Berrens, R. V., Andrews, S., Santos, F., Collier, A. J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P. J. and Reik, W. (2016). Comparative principles of DNA methylation reprogramming during human and mouse *in vitro* primordial germ cell specification. *Dev. Cell* **39**, 104-115. doi:10.1016/j.devcel.2016.09.015
- Wamaitha, S. E. and Niakan, K. K. (2018). Human pre-gastrulation development. *Curr. Top. Dev. Biol.* **128**, 295-338. doi:10.1016/bbs.ctdb.2017.11.004

- Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. and Brivanlou, A. H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **11**, 847-854. doi:10.1038/nmeth.3016
- Weber, S., Eckert, D., Nettersheim, D., Gillis, A. J. M., Schäfer, S., Kuckenberger, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L. H. J. et al. (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biol. Reprod.* **82**, 214-223. doi:10.1095/biolreprod.109.078717
- Wislocki, G. B. and Bennett, H. S. (1943). The histology and cytology of the human and monkey placenta, with special reference to the trophoblast. *Am. J. Anat.* **73**, 335-449. doi:10.1002/aja.1000730303
- Witschi, E. (1948). *Migration of the Germ Cells of Human Embryos from the Yolk Sac to the Primitive Gonadal Folds*. 575th edn, Carnegie Institution of Washington publication.
- Wu, J., Okamura, D., Li, M., Suzuki, K., Luo, C., Ma, L., He, Y., Li, Z., Benner, C., Tamura, I. et al. (2015). An alternative pluripotent state confers interspecies chimaeric competency. *Nature* **521**, 316-321. doi:10.1038/nature14413
- Yamaguchi, Y. and Yamada, S. (2018). The kyoto collection of human embryos and fetuses: history and recent advancements in modern methods. *Cells Tissues Organs* **205**, 314-319. doi:10.1159/000490672
- Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y. and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat. Genet.* **40**, 1016-1022. doi:10.1038/ng.186
- Ying, Y. and Zhao, G.-Q. (2001). Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev. Biol.* **232**, 484-492. doi:10.1006/dbio.2001.0173
- Ying, Y., Liu, X.-M., Marble, A., Lawson, K. A. and Zhao, G.-Q. (2000). Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol. Endocrinol.* **14**, 1053-1063. doi:10.1210/mend.14.7.0479
- Zhao, G.-Q. (2003). Consequences of knocking out BMP signaling in the mouse. *Genesis* **35**, 43-56. doi:10.1002/gene.10167
- Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S. N., Li, Z., Muncie, J. M., Lakins, J. N., Weaver, V. M., Gumucio, D. L. et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573**, 421-425. doi:10.1038/s41586-019-1535-2
- Zhu, Q., Sang, F., Withey, S., Tang, W., Dietmann, S., Klisch, D., Ramos-Ibeas, P., Zhang, H., Requena, C. E., Hajkova, P. et al. (2020). Specification and epigenetic resetting of the pig germline exhibit conservation with the human lineage. *Cell Rep.* **34**, 108735. doi:10.1016/j.celrep.2021.108735

Chapter 3

Germline competency of human embryonic stem cells depends on eomesodermin

Research Article

Germline competency of human embryonic stem cells depends on eomesodermin[†]

Di Chen¹, Wanlu Liu², Anastasia Lukianchikov¹, Grace V. Hancock^{1,2},
Jill Zimmerman¹, Matthew G. Lowe^{1,2}, Rachel Kim³, Zoran Galic^{3,4},
Naoko Irie^{5,6,7}, M. Azim Surani^{5,6,7}, Steven E. Jacobsen^{1,2,3,8,9}
and Amander T. Clark^{1,2,3,10,*}

¹Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, California, USA; ²Molecular Biology Institute, University of California, Los Angeles, California, USA; ³Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, California, USA; ⁴Department of Medicine, University of California, Los Angeles, California, USA; ⁵Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK; ⁶Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK; ⁷Wellcome Trust-Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, UK; ⁸Department of Biological Chemistry, University of California, Los Angeles, California, USA; ⁹Howard Hughes Medical Institute, University of California, Los Angeles, California, USA and ¹⁰Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California, USA

*Correspondence: Department of Molecular Cell and Developmental Biology, 615 Charles E Young Drive South, University of California Los Angeles, Los Angeles, CA 90095, USA. E-mail: clarka@ucla.edu

[†]Grant Support: All experiments, with the exception of human hESC derivation, were funded by R01 HD079546 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) (ATC) and supported by Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research NIH-NCATS UCLA CTSI Grant Number UL1TR0001881. No NIH funds were used for human embryo culture or hESC derivation. Instead, the derivation of hESC lines was funded by the California Institute for Regenerative Medicine, the UCLA Eli, and Edythe Broad Center of Regenerative Medicine and Stem Cell Research (BSCRC). Di Chen is supported by a training grant from the UCLA BSCRC. Wanlu Liu is supported by Philip J. Whitcome fellowship from the UCLA Molecular Biology Institute and a scholarship from the Chinese Scholarship Council. Human fetal tissue was obtained from the Laboratory of Developmental Biology, University of Washington, Seattle, which is supported by NIH Award Number 5R24HD000836 from the NICHD.

Received 23 October 2017; Accepted 27 October 2017

Abstract

In humans, germline competency and the specification of primordial germ cells (PGCs) are thought to occur in a restricted developmental window during early embryogenesis. Despite the importance of specifying the appropriate number of PGCs for human reproduction, the molecular mechanisms governing PGC formation remain largely unexplored. Here, we compared PGC-like cell (PGCLC) differentiation from 18 independently derived human embryonic stem cell (hESC) lines, and discovered that the expression of primitive streak genes were positively associated with hESC germline competency. Furthermore, we show that chemical inhibition of TGF β and WNT signaling, which are required for primitive streak formation and CRISPR/Cas9 deletion of Eomesodermin (*EOMES*), significantly impacts PGCLC differentiation from hESCs. Taken together, our results suggest that human PGC formation involves signaling and transcriptional programs associated with somatic germ layer induction and expression of *EOMES*.

Summary Sentence

EOMES induction in the progenitor cell prior to germ cell formation in vitro from hESCs is required for efficient PGC-like cell formation.

Key words: human, EOMES, primordial germ cells, embryonic stem cells.

Introduction

Primordial germ cells (PGCs) are diploid embryonic progenitor cells that ultimately differentiate into haploid gametes. PGCs are specified, maintained, and differentiated in a step-wise manner, and as a result are responsible for the number and quality of adult gametes. Very little is known about the earliest steps in human PGC differentiation in the embryo. Most of our knowledge comes from the mouse [1, 2], and more recently the nonhuman primate cynomolgus (cyno) macaque [3], rhesus macaque [4], and porcine [5]. Analysis of PGCs from cyno and porcine embryos relative to the mouse strongly suggest that the molecular and cellular events in PGC specification are different. Notably, around the time of PGC specification, cyno and porcine embryos develop morphologically as a bilaminar disk, whereas mouse embryos develop as an egg cylinder. In mouse embryos, PR domain zinc finger protein 1 (PRDM1), also called BLIMP1, transcription factor AP2 gamma (TFAP2C), and PR domain zinc finger protein 14 (PRDM14) constitute the critical tripartite transcription factor network responsible for specifying PGC fate from competent epiblast cells in vivo [6–8] and epiblast-like cells in vitro [9, 10]. In contrast, PRDM14 may not be required for human PGC fate [11], and instead, SOX17 has emerged as a new critical regulator in primates [12]. Therefore, the transcriptional network required for PGC specification within the mammalian class appears to have diversified.

Just as the transcriptional network for PGC specification has changed in different mammalian species, the timing of PGC formation in the peri-implantation embryo may have also diverged. In cyno embryos, PGCs are first identified in the peri-implantation embryo prior to primitive streak formation in an extra embryonic cell layer called the amnion [3]. Notably, the amnion is derived from the inner cell mass/epiblast just prior to primitive streak formation. In porcine embryos, PGC specification begins in the proximal pre-streak epiblast [5], whereas in the mouse embryo, PGC precursors are first identified as a small cluster in the proximal epiblast around the time of primitive streak formation [6, 13, 14]. Therefore, in cyno, mouse, and porcine embryos, where PGC specification has been studied in detail, PGC formation is induced from embryonic progenitors at or just prior to formation of the primitive streak and would therefore be exposed to signaling events that establish the primary germ layers and the primitive streak.

One of the earliest markers of primitive streak is the T-box transcription factor BRACHYURY (T), and in mouse embryos, T functions downstream of Wnt signaling to promote PGC specification. [15]. A second T-box transcription factor comesodermin (EOMES) is also expressed in the proximal epiblast prior to overt primitive streak formation, becoming restricted to the primitive streak during gastrulation [16]. In the mouse, there is no functional evidence for EOMES in the regulation of PGC development. However, using human-induced pluripotent stem cells (hiPSCs) a recent study indicated that EOMES is required for human PGC-like cell (PGCLC) formation following induced reprogramming [17]. It is not known whether EOMES is required for differentiation of human PGCLCs from human embryonic stem cells (hESCs).

Although transcription factors regulating PGC specification may have diverged between mice and humans [12], the major signaling pathways that specify PGCs have not. Most notably, bone morphogenetic protein 4 (BMP4) is essential for mouse PGC formation in vivo [18] as well as human and cyno PGCLC formation in vitro [5, 11, 12, 19, 20]. In the mouse, BMP4 specifies PGCs from the Wnt3-primed posterior epiblast, with T functioning downstream of WNT to regulate PGC fate [15]. In cyno's, T is expressed in the nucleus of nascent PGCs in the amnion, presumably due to Wnt3A signaling [3], and in porcine embryos, T is expressed in nascent PGCs of the preprimitive-streak embryo. Furthermore, porcine PGC numbers diminish following WNT inhibition [5]. These results suggest a conserved role for Wnt and BMP signaling in mouse, porcine, cyno, and human PGC formation in the embryo.

Human PGCLCs induced from multiple hiPSCs have been used to study the germ cell fate determinants [17, 21]. However, much less is known about the competency across different hESCs. In the current study, we discovered a positive correlation between the expression of primitive streak genes and efficiency of human PGCLC induction from 18 independently derived hESC lines. Using CRISPR/Cas9, we provide direct evidence that the transcription factor EOMES is required for efficient induction of human PGCLCs from hESCs, and lends additional support to the hypothesis that human PGCs are induced from cells in the embryo that express EOMES.

Material and methods

Human fetal samples

Human prenatal testes and ovaries were acquired following elected termination and pathological evaluation after University of California, Los Angeles (UCLA)-Institutional Review Board (IRB) review, which deemed the project exempt under 45 CFR 46.102(f). All prenatal gonads 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 collected the fetal testes and ovaries and shipped them overnight in HBSS with ice pack for immediate processing in Los Angeles. All consented material was donated anonymously and carried no personal identifiers. Developmental age was documented by BDRL as days post fertilization using a combination of prenatal intakes, foot length, Streeter Stages, and crown-rump length. All prenatal gonads documented with birth defect or chromosomal abnormality, were excluded from this study.

hESC culture

All primed hESC lines were cultured on mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) in hESC media, which is composed of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100 μ M L-Glutamine (GIBCO, 25030-081), 1 \times MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 55 μ M 2-Mercaptoethanol (GIBCO, 21985-023), 10 ng/mL recombinant

human FGF basic (R&D systems, 233-FB), 1× Penicillin-Streptomycin (GIBCO, 15140-122), and 50 ng/mL primocin (InvivoGen, ant-pm-2) in DMEM/F12 media (GIBCO, 11330-032). All hESC lines were split every 7 days with Collagenase type IV (GIBCO, 17104-019). 4i hESCs were maintained as described before [19]. 4i cells were grown on irradiated MEFs (GlobalStem) in knockout DMEM containing 20% KSR, 2 mM L-glutamine, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol (all GIBCO), 20 ng/mL human LIF (Stem Cell Institute [SCI]), 8 ng/mL bFGF (SCI), 1 ng/mL TGF- β 1 (Peprotech), 3 μ M CHIR99021 (Miltenyi Biotec), 1 μ M PD0325901 (Miltenyi Biotec), 5 μ M SB203580 (TOCRIS bioscience), and 5 μ M SP600125 (TOCRIS bioscience). 4i hESCs were split every 3 to 5 days using TrypLE Express (GIBCO). An amount of 10 μ M of ROCK inhibitor (Y-27632, TOCRIS bioscience) was used for the first 24 h after passage. All hESC lines used in this study are registered with the National Institute of Health Human Embryonic Stem Cell Registry and are available for research use with NIH funds. Specifically, the following hESC lines were used in this study: UCLA1 (46XX), UCLA2 (46XY), UCLA3 (46XX), UCLA4 (46XX), UCLA5 (46XX), UCLA6 (46XY), UCLA7 (47XX+13), UCLA8 (46XX), UCLA9 (46XX), UCLA10 (46XY), UCLA11 (46XY), UCLA12 (46XX), UCLA13 (46XY), UCLA14 (46XX), UCLA15 (46XX), UCLA16 (46XX), UCLA17 (46XX), UCLA18 (46XX). The derivation and basic characterization (Karyotype and teratoma analysis) of UCLA1–6 were previously reported [22]. UCLA8–10, UCLA14, and UCLA16–18 were reported [23].

Derivation and characterization of ESC lines from human embryos

The following UCLA hESC lines UCLA7 (47XX+13), UCLA11 (46XY), UCLA12 (46XX), UCLA13 (46XY), UCLA15 (46XX) were derived from human embryos according to the methods described [22]. All hESC derivations were performed after human subjects' approval from the UCLA-IRB and following Embryonic Stem Cell Research Oversight committee approval. No NIH funds were used for the derivation or initial characterization (karyotype and teratoma) of hESC lines. Teratoma analysis was performed after Institutional approval by the UCLA Office of Animal Research Oversight. Teratomas were created by injecting of collagenase digested clumps of hESCs into the testicles of male scid/beige C.B.17-Prkdc(scid)Lyst(bg) mice. Prior to injection, hESCs were resuspended in ice-cold matrigel (Corning, 354277), with 3× wells (from a 6-well plate) of colonies injected per testis. Karyotype analysis was conducted by Cell Line Genetics.

Induction of human PGCLCs though incipient mesoderm-like cells from primed hESCs

Human PGCLCs were induced from primed hESCs as described in [19] with some modifications. Day 7 hESCs were dissociated into single cells with 0.05% Trypsin-EDTA (Gibco, 25300-054) and plated onto Human Plasma Fibronectin (Invitrogen, 33016-015)-coated 12-well plate at the density of 200 000 cells/well in 2 mL/well of incipient mesoderm-like cell (iMeLC) media, which is composed of 15% KSR, 1× NEAA, 0.1 mM 2-Mercaptoethanol, 1× Penicillin-Streptomycin-Glutamine (Gibco, 10378-016), 1 mM sodium pyruvate (Gibco, 11360-070), 50 ng/mL Activin A (Peprotech, AF-120-14E), 3 μ M CHIR99021 (Stemgent, 04-0004), 10 μ M of ROCKi (Y27632, Stemgent, 04-0012-10), and 50 ng/mL primocin in Glasgow MEM (GMEM) (Gibco, 11710-035). iMeLCs were dissociated into single cells with 0.05% Trypsin-EDTA after 24 h of

incubation unless otherwise mentioned and plated into ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at the density of 3000 cells/well in 200 μ L/well of PGCLC media, which is composed of 15% KSR, 1× NEAA, 0.1 mM 2-Mercaptoethanol, 1× Penicillin-Streptomycin-Glutamine (Gibco, 10378-016), 1 mM sodium pyruvate (Gibco, 11360-070), 10 ng/mL human LIF (Millipore, LIF1005), 200 ng/mL human BMP4 (R&D systems, 314-BP), 50 ng/mL human EGF (R&D systems, 236-EG), 10 μ M of ROCKi (Y27632, Stemgent, 04-0012-10), and 50 ng/mL primocin in GMEM (Gibco, 11710-035). An amount of 100 ng/mL stem cell factor (SCF; PEPROTECH, 250-03), 10 μ M SB431542 (Stemgent, 04-0010-10), and 500 ng/ μ L Dickkopf Wnt Signaling Pathway Inhibitor 1 (DKK1) (R&D systems, 5439-DK) was added in PGCLC media for some experiments.

Induction of human PGCLCs from 4i hESCs

4i WIS2 cells were dissociated with TrypLE and plated to ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at the density of 2000–4000 cells/well in 200 μ L PGCLC media, which is composed of 15% KSR, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol, 100 U/mL Penicillin–0.1 mg/mL Streptomycin, 2 mM L-L-Glutamine, 1 mM Sodium pyruvate, 500 ng/mL BMP4 (R&D Systems) or BMP2 (SCI), 1 μ g/mL human LIF (SCI), 100 ng/mL SCF (R&D Systems), 50 ng/mL EGF (R&D Systems), and 10 μ M ROCK inhibitor in GMEM (Gibco, 11710-035).

Flow cytometry and fluorescence activated cell sorting

Human prenatal gonads or day 4 aggregates were dissociated with 0.25% trypsin (Gibco, 25200-056) for 5 min or 0.05% Trypsin-EDTA (Gibco, 25300-054) for 10 min at 37°C. The dissociated cells were stained with conjugated antibodies, washed with fluorescence activated cell sorting (FACS) buffer (1% BSA in PBS), and resuspended in FACS buffer accompanying with 7-AAD (BD Pharmingen, 559925). The single cell suspension was analyzed or sorted. The conjugated antibodies used in this study are ITGA6 conjugated with BV421 (BioLegend, 313624), EPCAM conjugated with 488 (BioLegend, 324210), EPCAM conjugated with APC (BioLegend, 324208), tissue nonspecific alkaline phosphatase (TNAP) conjugated with PE (BD Pharmingen, 561433), and cKIT conjugated with APC (BD Pharmingen, 550412).

Real-time quantitative polymerase chain reaction

Sorted cells or cell pellets were lysed in 350 μ L of RLT buffer (QIAGEN) and RNA was extracted using RNeasy micro kit (QIAGEN, 74004). Complementary DNA was synthesized using SuperScript II Reverse Transcriptase (Invitrogen, 18064-014). Real-time quantitative polymerase chain reaction (PCR) was performed using TaqMan Universal PCR Master Mix (Applied Biosystems, 4304437) and the expression level of genes-of-interest were normalized to the expression of housekeeping gene GAPDH. The Taqman probes used in this study include GAPDH (Applied Biosystems, Hs99999905.m1), NANOS3 (Applied Biosystems, Hs00928455.s1), PRDM1 (Applied Biosystems, hs01068508.m1), TFAP2C (Applied Biosystems, Hs00231476.m1), SOX17 (Applied Biosystems, Hs00751752.s1), cKIT (Applied Biosystems, hs00174029.m1), DAZL (Applied Biosystems, hs00154706.m1), DDX4 (Applied Biosystems, Hs00251859.m1), SOX9 (Applied Biosystems, Hs01001343.g1), AMH (Applied Biosystems, Hs01006984.g1), DND1 (Applied Biosystems, Hs00832091.s1), T (Applied Biosystems, Hs00610080.m1), CER1 (Applied Biosystems,

Hs00193796.m1), *MIXL1* (Applied Biosystems, Hs00430824.g1), *WNT3* (Applied Biosystems, Hs00902257.m1), *EOMES* (Applied Biosystems, Hs00172872.m1), *GSC* (Applied Biosystems, Hs00906630.g1), *FOXA2* (Applied Biosystems, Hs00232764.m1), *POU3F1* (Applied Biosystems, Hs00538614.s1), *TCF15* (Applied Biosystems, Hs00231821.m1). For human gonad samples, two technical replicates of real-time quantitative PCR were performed. For hESCs, iMeLCs, and PGCLCs, two to three independent experiments were performed.

Generation of *EOMES* mutant hESC lines

A pair of guide RNA (gRNA) targeting *EOMES* was designed using crispr.mit.edu, and the corresponding gDNA sequence was cloned into px330 vector [24]. An amount of 4 μ g of gRNA pair or 2 μ g of pMax-GFP was electroporated into 800 000 UCLA1 cells using P3 Primary Cell 4D-Nucleofector X Kit according to the manufacturer's instructions (Lonza, V4XP-3024). Twenty-four hours after nucleofection, cells were dissociated with Accutase (ThermoFisher Scientific, A1110501) and seeded in low density. A total of 96 individual colonies were picked after 9 days and expanded. Lines were screened for the presence of shorter bands due to deletion. To determine the precise mutations, PCR product from the targeted allele was cloned using Topo-TA cloning (Thermo-Fisher) and analyzed by Sanger sequencing. Two mutant lines were chosen and subcloned before experiments. The gRNA sequences used to target *EOMES* are 5'-GCGGTGTACAGCCGTAACAT and 5'-GTTATCTACACCGAAAGTGC. Karyotyping by Cell Line Genetics was performed before experimentation. GFP-labeled control and *EOMES* mutant hESCs were made by lentivirus-based transfection of UbiC-GFP-IRES-Puromycin and maintained as stable cell lines with puromycin (1 μ g/mL) selection.

Immunofluorescence

Immunostaining paraffin sections or aggregates in whole mount were described previously [25, 26]. For cells cultured on chamber slides, samples were fixed in 4% paraformaldehyde in PBS for 10 min and washed with PBS containing 0.1% Tween 20 and permeabilized with PBS containing Triton X for 10 min. Samples were blocked with 10% donkey serum for 30 min before antibody incubation. The primary antibodies used for immunofluorescence in this study include rabbit-anti-EPCAM (Abcam, ab71916, 1:50), goat-anti-VASA (R&D Systems, AF2030, 1:20), rabbit-anti-cKIT (DAKO, A4502, 1:100), goat-anti-OCT4 (Santa Cruz Biotechnology, sc-8628x, 1:100), rabbit-anti-PRDM1 (Cell Signaling Technology, 9115, 1:100), mouse-anti-PRDM1 (R&D Systems, MAB36081SP, 1:100), rabbit-anti-TFAP2C (Santa Cruz Biotechnology, sc-8977, 1:100), mouse-anti-TFAP2C (Santa Cruz Biotechnology, sc12762, 1:100), rat-anti-ITGA6 (Santa Cruz Biotechnology, sc-80554, 1:100), goat-anti-T (R&D Systems, AF2085, 1:100), goat-anti-SOX17 (Neuromics, GT15094, 1:100), rabbit-anti-EOMES (Abcam, ab23345, 1:100), rabbit-anti- β -CATENIN (Cell Signaling Technology, 9582, 1:100), rabbit-anti-pSMAD2/3 (Cell Signaling Technology, 8828, 1:100). The secondary antibodies used in this study are all from Jackson ImmunoResearch Laboratories including donkey-anti-rabbit-488, donkey-anti-mouse-488, donkey-anti-goat-488, donkey-anti-rat-488, donkey-anti-rabbit-594, donkey-anti-mouse-594, and donkey-anti-goat-594. DAPI is counterstained to indicate nuclei.

RNA-sequencing

Sorted cells or cell pellets were lysed in 350 μ L of RLT buffer (QIAGEN), and total RNA was extracted with RNeasy micro kit (QIAGEN, 74004). Total RNA was reverse transcribed and cDNA was amplified using Ovation RNA-Seq System V2 (Nugen, 7102-32) according to manufacturer's instructions. Amplified cDNA was fragmented into ~200 bp by Covaris S220 Focused-ultrasonicators. RNA-sequencing (RNA-seq) libraries were generated using Ovation Rapid Library Systems (Nugen, 0319-32 for index 1-8 and 0320-32 for index 9-16) and quantified by KAPA library quantification kit (Illumina, KK4824). Libraries were subjected to single-end 50 bp sequencing on HiSeq 2000 or HiSeq 2500 sequencer with 4-6 indexed libraries per lane.

RNAseq analysis

Analysis of individual gene expression

Raw reads in qseq format obtained from sequencer were first converted to fastq format with customized perl script. Reads quality were controlled with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). High-quality reads were then aligned to hg19 reference genome using Tophat [27] (v 2.0.13) by using "-no-coverage-search" option, allowing up to two mismatches and only keeping reads that mapped to one location. Basically, reads were first mapped to hg19 gene annotation with known splice junction. When reads did not map to the annotated genes, the reads were mapped to hg19 genome. Number of reads mapping to genes were calculated by HTseq [28] (v 0.5.4) with default parameters. Expression levels were determined by RPKM (reads per kilobase of exons per million aligned reads) in R using customized scripts. For RNAseq of published datasets GSE60138 [12], GSM1643143 [19], raw reads were obtained from GEO and then processed exactly the same as described above.

Hierarchical clustering of RNAseq

Raw read counts for each gene obtained from HTseq were preprocessed with DESeq R package [28]. To account for heteroscedasticity between samples, variance stabilizing transformation was first applied to all genes with DESeq. Samples were then hierarchical clustered (hclust function) based on their Euclidian distances (dist function) in R using customized scripts.

Principal component analysis

For principal component analysis (PCA), RPKM for each sample was first calculated. Variance of each genes across samples were then calculated (rowVars function in R). PCA analysis (prcomp function in R) was performed on genes with the top 500 variation across samples. PCA were then plotted with ggplot2 package in R (<http://ggplot2.org>).

Correlation analysis between samples

Correlation between ITGA6/EPCAM and TNAP/cKIT was calculated and plotted in R. Firstly, RPKM of each ITGA6/EPCAM and TNAP/cKIT RNAseq replicates were calculated. Pearson correlations were then obtained with cor function. Scatter plot was plotted with ggplot2 with linear regression line. Human PGC-enriched genes compared to H9 hESCs were defined by Irie et al. [12] (only genes with at least four fold enrichment in hPGC compared to H9 hESCs were selected). The RPKM of those selected genes were extracted, and heat map was plotted with log2(RPKM+1) values in R using pheatmap package.

Differential gene expression calling

R DESeq package was used to normalize counts per RefSeq transcripts to evaluate differential expression. For comparison between hESC, iMeLC, mutant hESC, and mutant iMeLC, highly upregulated genes of each sample, with mean $\log_2(\text{fold change}) > 1$ and adjusted P value < 0.05 were selected and plotted as Venn diagram using VennDiagram package in R. Scatter plot of $\log_2(\text{RPKM}+1)$ was plotted in R with ggplot2 package. $\log_2(\text{Fold change})$ were colored with different scale.

Analysis of gene expression level and PGCLC induction efficiency

In order to obtain iMeLC-specific genes coexpressing with PGCLC induction efficiency, Spearman correlation of each gene's expression level in iMeLC was calculated with PGCLC induction efficiency. Genes in hESC with RPKM more than 2 were filtered out, while genes with at least 2 RPKM in iMeLC were kept. Gene's expression levels with at least 0.45 correlations to PGCLC induction efficiency were kept. Heat map of $\log_2(\text{RPKM}+1)$ was plotted in R as described above.

Results

ITGA6 and EPCAM can be used to isolate human PGCs from embryonic ovaries but not embryonic testes

Recently, *in vitro* human PGCLCs were isolated using INTEGRIN α 6 (ITGA6) and EPCAM following hiPSC differentiation [19]; however, whether ITGA6/EPCAM can be used to isolate *in vivo* PGCs from human embryos has not been shown. To address this, we examined cells from a pair of human embryonic ovaries at 72 day postfertilization by staining with four antibodies that detect ITGA6, EPCAM, TNAP, and cKIT. cKIT and TNAP were used as a positive control, given that these surface proteins can be used to sort human PGCs from the prenatal embryonic ovary and testis [25, 29]. The labeled cells were divided in two and sorted by FACS gating on either ITGA6/EPCAM or TNAP/cKIT (Figure 1A). We discovered that the percentage of ITGA6/EPCAM double positive cells and TNAP/cKIT double positive cells in the embryonic ovary was comparable (Figure 1A). Furthermore, the ITGA6/EPCAM double positive population was also double positive for TNAP/cKIT. Conversely, the TNAP/cKIT double positive cells were also double positive for ITGA6/EPCAM (Figure 1A). Using real-time PCR, we show that both populations expressed PGC genes at equivalent levels (Figure 1B). We repeated this experiment with a pair of human embryonic ovaries at 94 day, and found a similar result (Figure 1A and B). Therefore, these observations indicate that ITGA6/EPCAM can be used to isolate TNAP/cKIT positive germ cells from the human embryonic ovary and vice versa.

Next, we performed the same analysis using embryonic testes. Unlike the ovary, we discovered that most ITGA6/EPCAM double positive cells are negative for TNAP/cKIT (Figure 1C), and the ITGA6/EPCAM double positive cells have reduced levels of PGC genes relative to the TNAP/cKIT double positive population (Figure 1D). Given that ITGA6 and EPCAM are epithelial markers [30, 31], and PGCs in the embryonic testis are known to be enclosed within epithelial cords [25], we reasoned that ITGA6 and EPCAM double positive cells are a mixture of both germ cells and somatic cells in prenatal testis. To test this, we performed immunofluorescence and real-time PCR and show that ITGA6 and EPCAM are expressed by both PGCs and epithelial Sertoli cells (Figure S1A and B), and the ITGA6/EPCAM double positive population is also en-

riched in Sertoli cell genes *SOX9* and *AMH* (Figure 1D). Therefore, using prenatal tissues containing PGCs, TNAP/cKIT can be used to sort PGCs from both the embryonic ovary and testis, whereas ITGA6/EPCAM can only be used to sort PGCs from the embryonic ovary.

Next, we performed RNA-seq on sorted ITGA6/EPCAM female PGCs at 89 and 103 day (Figure 1E and F), and compared this to a third sample involving a pair of ovaries collected at 89 d, which were pooled and then stained and sorted separately for ITGA6/EPCAM or TNAP/cKIT (Figure 1E–G, marked by asterisk). We also included a testis at 59 day, which was sorted only using TNAP/cKIT to specifically isolate the male PGCs. Using unsupervised hierarchical clustering (UHC), the sorted PGC samples clustered together forming a distinct group relative to an undifferentiated female hESC line (UCLA1), and a male hESC line (UCLA2) (Figure 1E). Evaluation of candidate genes revealed that all putative PGC samples expressed the early and late stage PGC genes and have acquired a naive pluripotent signature (Figure S1C). Moreover, the ITGA6/EPCAM and TNAP/cKIT PGCs sorted from the same pair of 89 day embryonic ovaries clustered the closest compared to ITGA6/EPCAM populations sorted from different ovaries (Figure 1E). This result further supports the conclusion that ITGA6/EPCAM and TNAP/cKIT are expressed on the same population of PGCs. To better display the similarities and differences between samples, we performed PCA, and found that the PGCs are well separated from undifferentiated hESCs along the PC1 axis, while the 59-day male TNAP/cKIT PGCs are separated from the other gonadal PGCs along the PC2 axis (Figure 1F). This separation is anticipated to be caused by the 4-week difference in age between male and female samples in this study, as well as gender [32]. Last, we examined the differentially expressed genes (DEGs) between ITGA6/EPCAM and TNAP/cKIT positive cells sorted from the same ovary pair and found very few DEGs as the two transcriptomes are highly correlated with linear regression R squared = 0.98 (Figure 1G). Therefore, we propose that ITGA6/EPCAM and TNAP/cKIT are coexpressed on the same cell population in the embryonic ovary.

Germline competency is associated with the induction of genes required for primitive streak formation

Prior to examining PGCLC differentiation across 18 independently derived UCLA hESC lines with ITGA6/EPCAM/TNAP/cKIT, we first piloted the sorting strategy on the male hESC line UCLA2. We discovered that all ITGA6/EPCAM double positive putative PGCLCs at day 4 of differentiation were positive for TNAP; however, cKIT was not detected on the PGCLC population (Figure S2A). In mice, cKIT is critical for PGC survival, migration, and proliferation [33–35]. In humans, cKIT is present on PGCs from at least week 4 to week 20 of prenatal life postfertilization [25, 29, 32, 36]. Furthermore, analysis of cyno PGCs during gastrulation reveals that *cKIT* RNA is expressed by PGCs at the time of specification [3]. Thus, we reasoned that cKIT is either not translated in early stage human PGCs, or alternatively the conditions for PGCLC induction must be optimized to enable cKIT expression on the cell surface. Given that cKIT is subject to ligand induced endocytosis [37], we removed the ligand SCF from the PGCLC media, and this resulted in ITGA6/EPCAM/TNAP/cKIT positive PGCLCs (Figure S2B and D). To determine if excluding SCF affects germ cell identity, we performed real-time PCR and RNA-Seq of the putative ITGA6/EPCAM positive PGCLCs differentiated with or without SCF (Figure S2C and E). To determine the molecular similarity of PGCLCs generated in primed conditions on MEFs, to those generated from hESCs cultured in 4i on MEFs [12], we

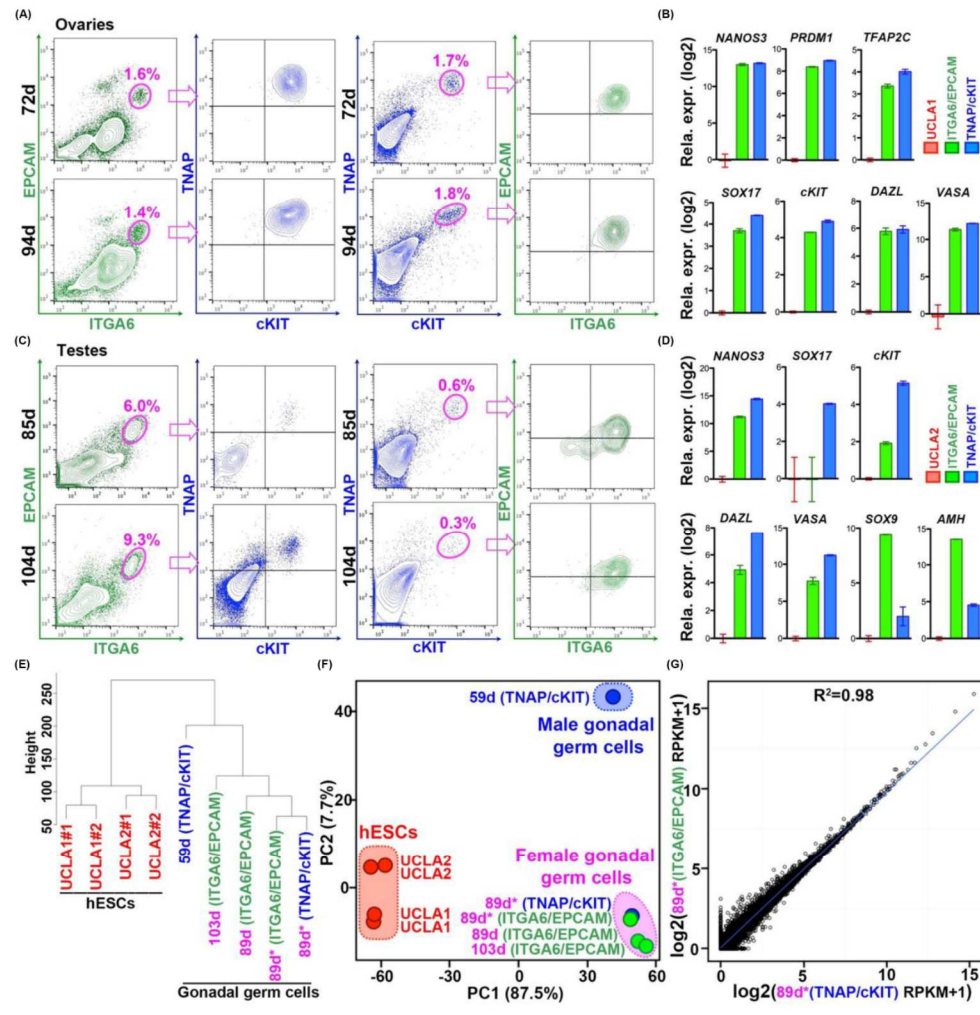


Figure 1. Analysis of ITGA6/EPCAM and TNAP/cKIT populations in human prenatal gonads. (A) Flow cytometry of prenatal ovaries at day (d) 72 and 94 postfertilization stained with antibodies that recognize ITGA6, EPCAM, TNAP, and cKIT. (B) Gene expression of the sorted populations from (A, 72d) by real-time PCR. Expression is normalized to the GAPDH. Fold change is calculated relative to expression levels of each gene in female hESC line UCLA1 (passage 17 (p17) and p18), which was given a value of 1.0. (C) Flow cytometry of prenatal testis at day 85 and 104 postfertilization stained with ITGA6, EPCAM, TNAP, and cKIT. (D) Gene expression of the sorted populations from (C, 85d) by real-time PCR. For each gene examined, its expression is normalized to the GAPDH. Fold change is calculated relative to expression levels of each gene in male hESC line UCLA2 (p11 and p12), which was given a value of 1.0. (E) Unsupervised hierarchical clustering (UHC) of transcriptomes of female hESC line UCLA1 (two biological replicates, p14 and p15), male hESC line UCLA2 (two biological replicates, p13 and p14), TNAP/cKIT positive germ cells from embryonic day 59 testes, ITGA6/EPCAM positive cells from 89d, 103d, and another 89d embryonic ovary. TNAP/cKIT positive cells from 89d ovaries. Asterisk in E–G indicates that the two RNA-seq libraries were made from the same pair of ovaries but sorted with different surface markers. (F) PCA of transcriptomes shown in E. (G) Scatter plot of two transcriptomes made from the same pair of ovaries but sorted with ITGA6/EPCAM (x-axis) and TNAP/cKIT (y-axis).

Downloaded from https://academic.oup.com/biolreprod/article/97/6/850/4582255 by guest on 21 May 2021

differentiated the WIS2 hESC line for 4 days according to the methods of Irie et al. [12] and sorted the TNAP/NANOS3-mCherry PGCLC population. RNA-seq analysis showed that PGCLCs generated from either the primed or 4i cultured hESCs clustered together, and were distinct from the undifferentiated hESCs (Figure S2E and F), indicating that the starting culture media (4i on MEFs versus primed media on MEFs) ultimately yielded PGCLCs with similar transcriptional identities.

Using ITGA6/EPCAM as a sorting approach to analyze PGCLCs, we next examined PGCLC competency of 18 hESC lines, all derived at UCLA from 18 single frozen embryos (Figure 2A). All hESC lines were capable of teratoma formation when injected into severe combined immunodeficient (SCID) beige mice (UCLA1–6 [22]; UCLA8–10, UCLA14, UCLA16–18 [23]; and this study UCLA7, UCLA11–13, and UCLA15; Figure S3A–C). Seventeen out of eighteen hESC lines are karyotypically normal, while UCLA7 has a duplication of chromosome 13 in 100% of cells karyotyped (Figure S3B). We included this cell line in the analysis to determine whether aneuploidy may be associated with alterations in PGCLC potential. It has been previously reported that PGCLC induction efficiency is variable between experiments [12, 19]. In order to minimize variability, we induced PGCLCs from 18 hESC lines simultaneously and repeated this experiment four times to determine the average PGCLC induction efficiency for all 18 lines. All 18 independently derived hESC lines were germline competent. However, UCLA6 consistently exhibited the highest germline potential generating on average 35% PGCLCs at day 4 of aggregate formation, whereas UCLA9 had the lowest germline potential generating on average less than 1% PGCLCs at day 4 (Figure 2A; Figure S3D). The aneuploid UCLA7 female hESC line generated a comparable percentage of PGCLCs to the other female hESC lines in the data set. Meanwhile, we found that male hESC lines on average were more competent for PGCLC induction than female (Figure 2B).

In a recent study using nine hiPSC lines, it was determined that primitive streak genes were associated with PGCLC competency [21]. To determine whether this was also the case for hESCs, we performed RNA-seq on the 18 hESC lines, and the corresponding iMeLCs in biological duplicate. We used the following parameters to define genes whose expression levels in iMeLCs are positively correlated with PGCLC induction efficiency. First, we chose genes that were expressed at low levels or not expressed in hESCs by setting the RPKM value < 2. Second, we used the RPKM value > 2 in at least one of the 18 iMeLC samples to select for genes that are upregulated in at least one of the 18 hESC lines induced to become iMeLCs. Third, we correlated gene expression in iMeLCs with PGCLC induction efficiency, and set the cutoff as >0.45. Based on these parameters, we found 78 genes upregulated from hESCs to iMeLCs that were also positively correlated with resulting PGCLC induction efficiency in the aggregates (Figure 2C). Using gene ontology (GO) term analysis of these 78 genes by WebGestalt [38, 39], we discovered that the top term was “formation of primary germ layer” (GO: 0001704), which included seven genes: *EOMES*, *T*, *GATA6*, *MIXL1*, *GATA4*, *WLS*, and *LHX1*. These genes are associated with primitive streak formation that are known to be induced by NODAL/ACTIVIN A (TGF β) and WNT [40].

Signaling pathways that promote primitive streak formation are also required for PGCLC induction

Next, we confirmed the relationship between TGF β and WNT signaling pathways, and the expression of *T* and *EOMES* in iMeLCs.

We performed immunofluorescence and identified nuclear accumulation of phosphorylated SMAD2/3 (pSMAD2/3) (Figure 3A) and β -CATENIN (Figure 3B) in both iMeLCs and hESCs, indicating that these cells are capable of responding to both signaling pathways. We then confirmed expression of *T* (Figure 3C) and *EOMES* (Figure 3D) protein by immunofluorescence, and as predicted from the RNA-Seq, *T* was induced in iMeLCs (Figure 3C), and *EOMES* protein was expressed in both hESCs and in iMeLCs (Figure 3D).

The correlation of germline competency with TGF β and WNT signaling, as well as *T* and *EOMES* induction promoted us to test if these molecules are critical for PGCLC formation. To block TGF β signaling we added SB431542 [41], which inhibits the TGF β type I receptors ALK5, ALK4, and ALK7, and phosphorylation of SMAD2 and SMAD3. This inhibitor does not block the ALKs or SMADs downstream of BMP4. To block WNT receptor binding, we added DKK1 [42] (Figure 3E). We discovered that the addition of these two molecules prevented the induction of *T* and *EOMES* in iMeLCs despite the presence of ACTIVIN A and GSK3 β in the iMeLC media (Figure 3F). In order to determine if *T* and *EOMES* expression is dependent on TGF β or WNT, or both, we evaluated *T* and *EOMES* expression in iMeLCs in the presence of either SB431542 or DKK1. We found that *T* and *EOMES* were repressed in the presence of either SB431542 or DKK1 (Figure S4A and B), suggesting the expression of *T* and *EOMES* in iMeLCs requires both TGF β and WNT signaling. To evaluate the effect on germline competency, we induced PGCLCs in media with or without SB431542 and DKK1 (Figure 3E). We discovered that the PGCLC population was completely abolished when SB431542 and DKK1 were included in the media, despite the presence of exogenous BMP4 and other cytokines known to induce PGCLC fate (Figure 3G). These results suggest that the ability to respond to TGF β and WNT signaling and potentially the induction of genes associated with primitive streak formation such as *T* and *EOMES* are required for germline induction.

EOMES is required for PGCLC induction

To provide direct evidence for *EOMES* in PGCLC induction, we used CRISPR/Cas9 to mutate *EOMES* in the UCLA1 hESC line (Figure S4C). We used two independent *EOMES* mutant clones for analysis. By differentiating the *EOMES* mutant and control hESC lines in parallel, we found that the ITGA6/EPCAM double positive PGCLC population was dramatically reduced in the *EOMES* mutant clones (Figure 4A–B). To confirm this result, we also performed immunofluorescence on aggregates and detected PGCLCs by triple staining for TFAP2C, SOX17, and PRDM1. In wild-type control cells, we detected clusters of triple TFAP2C, SOX17, and PRDM1 human PGCLCs (Figure 4C, white dot outlined). However, in the *EOMES* mutant lines, we identified very few triple positive PGCLCs (Figure 4C, white arrowheads). Therefore, our results demonstrate that *EOMES* is required for PGCLC formation in human, most likely downstream of WNT and TGF β .

In cyno embryos, nascent PGCs in the amnion are negative for *EOMES* [3]. Similarly, our RNA-seq data show that *EOMES* is not expressed in prenatal PGCs or PGCLCs. Therefore, we hypothesize that *EOMES* acts either noncell autonomously in the niche to support PGCLC formation, or alternatively *EOMES* acts cell autonomously in iMeLCs to prime PGC fate. To test this, we made stable GFP lines of control and *EOMES* mutant hESCs. We made iMeLCs and generated PGCLC aggregates containing a mix of GFP and unlabeled cells in a 1:1 ratio. The iMeLC mixtures were composed of GFP-labeled control cells with unlabeled controls (Figure 4D) or

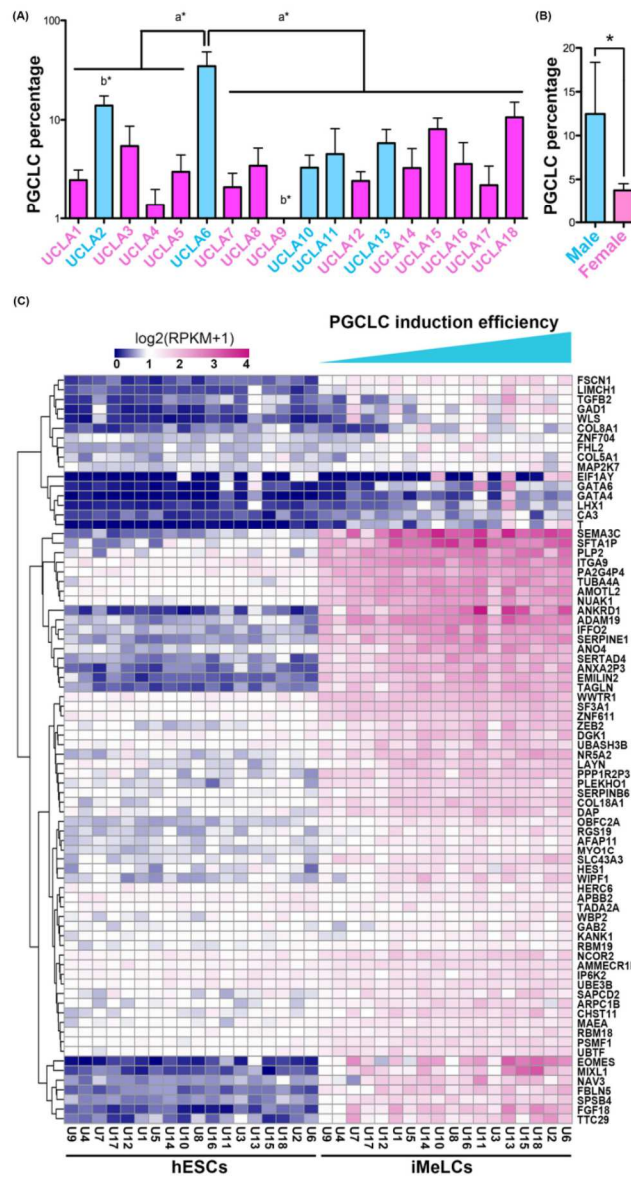


Figure 2. Germline competency varies between independent hESC lines. (A) Average PGCLC induction efficiency at day 4 in aggregates generated from 18 hESC lines (passage numbers ranging from p10 to p22, see experimental procedures for details). Blue represents male and pink represents female. "a*" indicates the significant difference between UCLA6 and all other cell lines and "b*" indicates the significant difference between UCLA2 and UCLA9 (tested by ANOVA, $P < 0.0001$). PGCLCs were identified as ITGA6/EPCAM double positive cells. (B) Comparison of day 4 PGCLC induction efficiency from male (blue) and female (pink) hESC lines. "*" indicates the difference between male and female (t-test, $P < 0.05$). (C) Heat map showing the expression of genes in iMeLCs that positively correlated with PGCLC induction efficiency. Genes are selected as maximal expression ≤ 2 (RPKM) in hESCs, maximal expression ≥ 2 (RPKM) in iMeLCs, and correlation coefficient > 0.45 .

Downloaded from <https://academic.oup.com/biolreprod/article/97/6/850/4582255> by guest on 21 May 2021

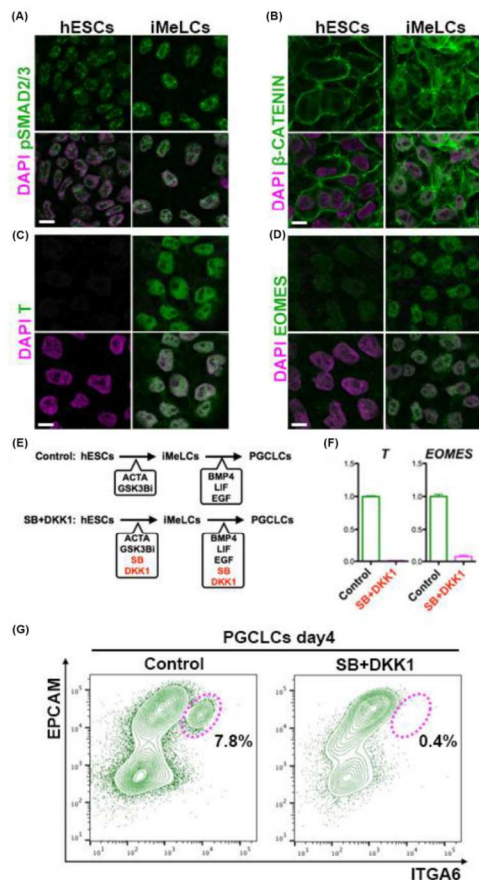


Figure 3. TGF β and WNT signaling are required for PGCLC induction from hESCs. (A–D) Expression of pSMAD2/3 (A), β -CATENIN (B), T (C), and EOMES (D) in UCLA1 hESCs and iMeLCs. Scale bar: 10 μ m. (E) Schematic illustration of PGCLC induction with or without SB431542 (SB) and DKK1. (F) Real-time PCR for *T* and *EOMES* expression at iMeLCs in the presence of SB431542 and DKK1. (G) Flow cytometry showing loss of PGCLC competency in media containing SB431542 and DKK1.

GFP-labeled *EOMES* mutant cells with unlabeled controls (Figure 4E). In the controls, about 45% of the total cells were positive for GFP, and about 35% of the PGCLCs were positive. Therefore, control hESCs with or without GFP both contribute to PGCLC induction (Figure 4D). In contrast, only about 8% of the PGCLCs were induced from GFP-labeled *EOMES* mutant cells, whereas the total cells were composed of about 48% GFP-labeled *EOMES* mutant cells (Figure 4E). This suggests that *EOMES* is required cell autonomously to prime PGC fate.

Discussion

In the current study, we prove that PGCLC competency is an inherent property of hESC lines. Together with previous work [11, 12, 19], our study proves that the generation of human PGCLCs in vitro through directed differentiation is not a stochastic event restricted to a small number of hESC or hiPSC lines, but rather a competency that extends to the majority of human pluripotent stem cell lines in vitro regardless of whether they originated from the inner cell mass or through induced reprogramming. We also show that heterogeneity in PGCLC differentiation can provide a unique opportunity to discover conserved and new genes that regulate PGC specification in humans. Heterogeneity in differentiation potential amongst independently derived pluripotent stem cell lines is not unique to the germline, with multiple studies showing that hESC and hiPSC lines each have varying potentials for somatic cell differentiation [43–46]. In these studies, the underlying cause of this heterogeneity is thought to be due to differences in chromatin and DNA methylation in the self-renewing state [45, 46]. The underlying cause of variable germline competency also warrants further investigation.

A recent study using nine hiPSC lines cultured on StemFit also revealed that genes associated with primitive streak formation are associated with germline competency of hiPSCs in vitro [21]. Our study extends this observation by demonstrating that primitive streak gene expression is also associated with germline competency of hESCs, and that functionally, PGCLC competency can be attributed to the appropriate induction of *T* and *EOMES* downstream of TGF β and WNT signaling. A recent study using hiPSCs revealed that *T* is not required for hPGCLC specification; therefore, even though *T* is induced and associated with PGCLC competency, *T* itself may not be necessary for germline competency in vitro [17].

Unlike *T*, *EOMES* has no reported functional role in mouse, cyno, or porcine PGC specification. In mouse, *EOMES* is required for the cell migration in the primitive streak and *EOMES* loss-of-function mutants fail to undergo germ layer formation [47]. In the current study, we determined that *EOMES* is critical for human PGCLC specification where *EOMES* functions cell autonomously. Our RNA-seq data show that *EOMES* is expressed in iMeLCs but not in PGCLCs or PGCs. Therefore, we hypothesize that *EOMES* functions before the specification of SOX17 positive PGCs and is associated with germ cell competency of the epiblast/primitive streak. Whether *EOMES* regulates the migration of iMeLCs and whether the migration of iMeLCs is required for germ cell fate specification require further analysis.

Finally, our work touches on the potential origins of the human germline. In mouse embryos, PRDM1 positive PGC precursors are specified by BMP4 signaling to the Wnt3 primed posterior epiblast around the time of primitive streak formation [6, 15, 48]. In cyno embryos, PGCs were first identified prior to primitive streak formation in an embryonic cell layer called the amnion [3]. In porcine embryos, SOX17 positive PGCs were identified in prestreak epiblasts similar to the mouse [5]. In the current study, we discovered that signaling pathways and transcription factors associated with primitive streak formation (*EOMES*) regulate competency for SOX17 positive human PGCLC formation. This does not directly refute the origin of PGCLCs as being from the amnion, as the molecular identity of amnion cells is almost completely unknown. Instead, our work demonstrates that embryonic cells with the appropriate competency to respond to TGF β and Wnt3 in order to induce *EOMES* are required for human PGCLC formation. Future studies using embryo attachment culture will be necessary to determine the origin of the

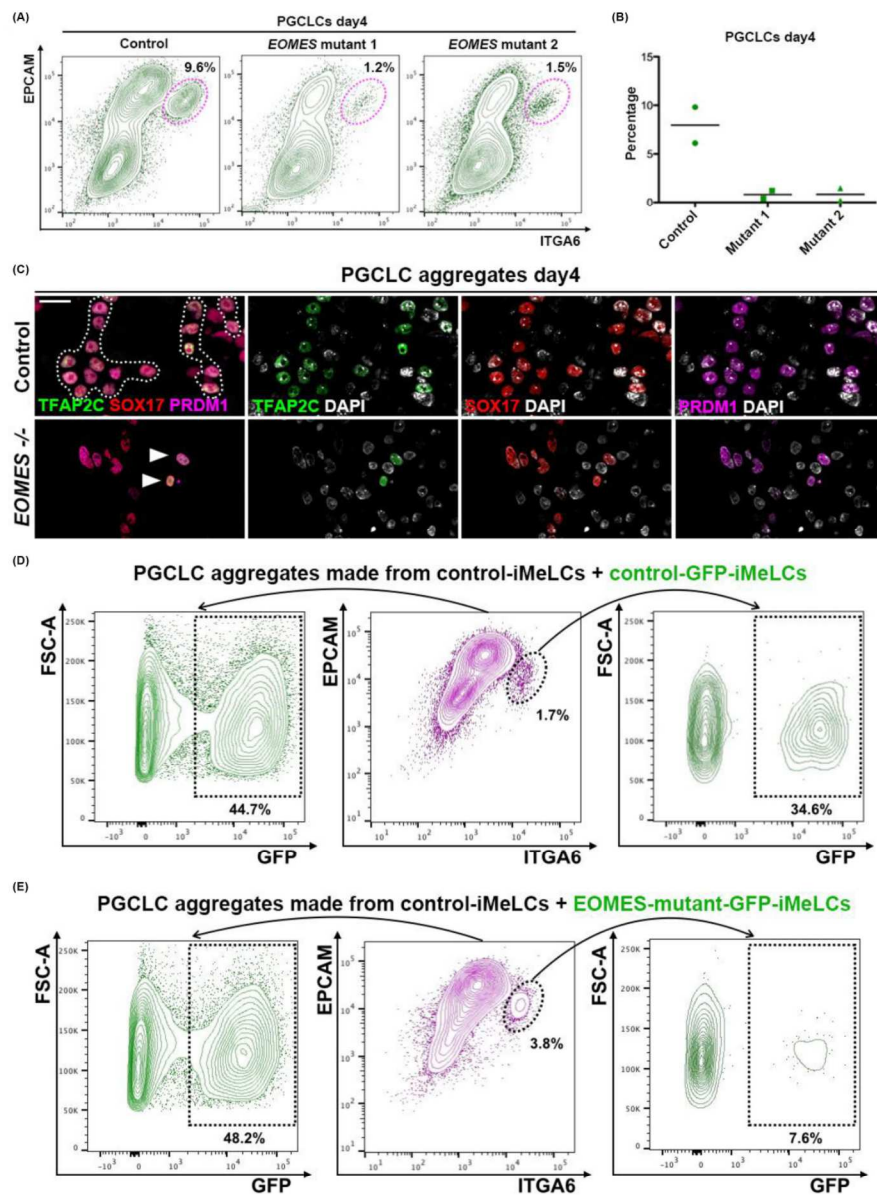


Figure 4. *EOMES* is required for PGCLC induction from hESCs. (A) Flow cytometry showing reduced PGCLC competency in *EOMES* mutant hESC line. PGCLCs were identified as ITGA6/EPCAM double positive cells. (B) Summary of PGCLC induction percentage from control and two different *EOMES* mutant hESC clones. (C) Control and *EOMES* mutant day 4 aggregates stained with TFAP2C (green), SOX17 (red), PRDM1 (purple), and DAPI (white). White dot outlines TFAP2C/SOX17/PRDM1 triple positive PGCLCs in control. White arrowheads point to rare TFAP2C/SOX17/PRDM1 triple positive PGCLCs in *EOMES* mutant. (D) Flow cytometry analysis of PGCLCs made from mixed iMeLCs (1:1 ratio) made from GFP negative and GFP positive wild-type hESCs. Left panel shows GFP positive cells in all live cells from PGCLC aggregates. Middle panel shows PGCLCs positive for ITGA6 and EPCAM. Right panel shows GFP positive PGCLCs. (E) Same analysis as (D) for PGCLCs made from mixed iMeLCs (1:1 ratio) made from GFP negative wild-type hESCs and GFP positive *EOMES* mutant hESCs.

human germline, but only if they emerge before the primitive streak given the ethical barrier that our field currently cannot cross.

Supplementary data

Supplementary data are available at BIOLOGYonline.

Figure S1. Germ cell marker expression and optimization of PGCLC induction from hESCs. (A) Human embryonic testis at day 115 section stained with ITGA6 (green), OCT4 (red), and DAPI (blue) by immunofluorescence. Yellow arrowhead points to an OCT4-positive germ cell that is also positive for ITGA6. Purple arrowhead points to an ITGA6-positive germ cell that is negative for OCT4. Scale bar: 10 μ m. (B) Human embryonic testis at day 72 section stained with EPCAM (green), VASA (red), and DAPI (blue) by immunofluorescence. Yellow arrowhead points to a VASA-positive germ cell that is also positive for EPCAM. Purple arrowhead points to an EPCAM-positive germ cell that is negative for VASA. Scale bar: 10 μ m. (C) Heat map showing the expression of germ cell, somatic cell, and pluripotency genes in transcriptomes shown in Figure 1E.

Figure S2. Generating ITGA6/EPCAM/TNAP/cKIT positive PGCLCs. (A and B) Flow cytometry of day 4 aggregates made with SCF (A) or without SCF (B) from UCLA2 (p13) and stained for ITGA6/EPCAM/TNAP/cKIT. (C) Real-time PCR of the sorted ITGA6/EPCAM positive PGCLCs made from UCLA2 (p13, p14) with SCF (green open columns) or without SCF (purple open columns) and compared to gene expression in ITGA6/EPCAM PGCs at day 72 postfertilization (Figure 1A and B) (green solid columns). Fold change is calculated relative to expression levels of each gene in the UCLA1 hESC line, which was given a value of 1.0. (D) Immunofluorescence of day 4 PGCLC aggregates from UCLA2 (p14) to examine colocalization of OCT4 (red), with cKIT, PRDM1, and TFAP2C (all green). Scale bar: 10 μ m. (E) UHC of primed undifferentiated hESCs (UCLA1 p14, p15 and UCLA2 p13, p14), day 4 PGCLCs sorted by FACS using ITGA6/EPCAM (made from UCLA1 p14, p15 and UCLA2 p13, p14) with (plus) and without (minus) SCF, undifferentiated 4i cultured hESCs sorted with TNAP (WIS2) and day 4 PGCLCs sorted by FACS using TNAP/NANOS3-mCherry (made from WIS2). UHC was based on the expression of DEGs between hPGCs and H9 hESCs defined by Irie et al. [12] and Sasaki et al. [19]. U1 and U2 indicate UCLA1 and UCLA2, respectively. Gonadal germ cell libraries analyzed here are the same in Figure 1E. MS: minus SCF. PS: plus SCF. (F) PCA of transcriptomes in (E).

Figure S3. PGCLC induction from 18 pluripotent hESC lines derived at UCLA. (A) Morphology of human embryos used for derivation of hESC lines UCLA7, UCLA11, UCLA12, UCLA13, and UCLA15. All other hESC lines are reported elsewhere. (B) Karyotypes of hESC lines UCLA7, UCLA11, UCLA12, UCLA13, and UCLA15 (from left to right). All other hESC lines are reported elsewhere. (C) Representative images showing teratomas formed by injection of hESC lines UCLA7, UCLA11, UCLA12, UCLA13, and UCLA15 (from left to right) into the testes of SCID-beige mice. All other hESC lines are reported elsewhere. (D) FACS plots of day 4 PGCLCs (sorted with ITGA6/EPCAM) induced from 18 hESC lines through 24 h of iMeLC differentiation.

Figure S4. Evaluation of *T* and *EOMES* in different combination of cytokines and inhibitors and molecular information of *EOMES* mutant alleles. (A) *T* expression in the iMeLCs with different combinations of cytokines and signaling inhibitors. (B) *EOMES* expression in the iMeLCs with different combinations of cytokines and signaling inhibitors. (C) Molecular information of gRNAs for targeting

EOMES and the resulting *EOMES* mutant alleles in the subline used in this study.

Table S1. List of antibodies used in this study.

Acknowledgments

The authors would like to thank Felicia Codrea and Jessica Scholes for FACS, Jinghua Tang for banking and culturing of the UCLA hESC lines, 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. Human conceptus tissue requests can be made to bdrl@u.washington.edu. SEJ is an investigator of the Howard Hughes Medical Institute.

Conflict of Interest: The authors have declared that no conflict of interest exists.

References

- Magnusdottir E, Surani MA. How to make a primordial germ cell. *Development* 2014; 141:245–252.
- Tang WWC, Kobayashi T, Irie N, Dietmann S, Surani MA. Specification and epigenetic programming of the human germ line. *Nat Rev Genet* 2016; 17:585–600.
- Sasaki K, Nakamura T, Okamoto I, Yabuta Y, Iwatani C, Tsuchiya H, Seita Y, Nakamura S, Shiraki N, Takakuwa T, Yamamoto T, Saitou M. The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Dev Cell* 2016; 39:169–185.
- Clark AT, Gkountela S, Chen D, Liu W, Sosa E, Sukhwani M, Hennebold JD, Orwig KE. Primate primordial germ cells acquire trans-plantation potential by Carnegie stage 23. *Stem Cell Reports* 2017; 9:329–341.
- Kobayashi T, Zhang H, Tang WWC, Irie N, Withey S, Klisch D, Sybirna A, Dietmann S, Contreras DA, Webb R, Allegrucci C, Alberio R et al. Principles of early human development and germ cell program from conserved model systems. *Nature* 2017; 546:416–420.
- Ohinata Y, Payer B, O'carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovskiy A, Saitou M, Surani MA. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 2005; 436:207–213.
- Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y, Saitou M. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* 2008; 40:1016–1022.
- Weber S, Eckert D, Nettersheim D, Gillis AJ, Schäfer S, Kuckenberg P, Ehlermann J, Werling U, Biermann K, Looijenga LH, Schorle H. Critical function of AP-2gamma/TFAP2C in mouse embryonic germ cell maintenance. *Biol Reprod* 2010; 82:214–223.
- Nakaki F, Hayashi K, Ohta H, Kurimoto K, Yabuta Y, Saitou M. Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* 2013; 501:222–226.
- Magnúsdóttir E, Dietmann S, Murakami K, Günesdogan U, Tang F, Bao S, Diamanti E, Lao K, Gottgens B, Azim Surani M. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat Cell Biol* 2013; 15:905–915.
- Sugawa F, Arauzo-Bravo MJ, Yoon J, Kim K-P, Aramaki S, Wu G, Stehling M, Psathaki OE, Hubner K, Scholer HR. Human primordial germ cell commitment in vitro associates with a unique PRDM14 expression profile. *EMBO J* 2015; 34:1009–1024.
- Irie N, Weinberger I, Tang WWC, Kobayashi T, Viukov S, Manor YS, Dietmann S, Hanna JH, Surani MA. SOX17 is a critical specifier of human primordial germ cell fate. *Cell* 2015; 160:253–268.
- Lawson KA, Hage WJ. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* 1994; 182:68–84; discussion 84–91.
- Tam PP, Zhou SX. The allocation of epiblast cells to ectodermal and germline lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* 1996; 178:124–132.

15. Aramaki S, Hayashi K, Kurimoto K, Ohta H, Yabuta Y, Iwanari H, Mochizuki Y, Hamakubo T, Kato Y, Shirahige K, Saitou M. A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev Cell* 2013; 27:516–529.
16. Ciruna BG, Rossant J. Expression of the T-box gene Eomesodermin during early mouse development. *Mech Dev* 1999; 81:199–203.
17. Kojima Y, Sasaki K, Yokobayashi S, Sakai Y, Nakamura T, Yabuta Y, Nakaki F, Nagaoka S, Wolftjen K, Hotta A, Yamamoto T, Saitou M. Evolutionarily distinctive transcriptional and signaling programs drive human germ cell lineage specification from pluripotent stem cells. *Cell Stem Cell* 2017; 21:517.e5–532.e5.
18. Lawson KA, Dunn NR, Roelen BAJ, Zeinstra LM, Davis AM, Wright CVE, Koryng JPWF, Hogan BLM. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* 1999; 13:424–436.
19. Sasaki K, Yokobayashi S, Nakamura T, Okamoto I, Yabuta Y, Kurimoto K, Ohta H, Moritoki Y, Iwatani C, Tsuchiya H, Nakamura S, Sekiguchi K et al. Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* 2015; 17:178–194.
20. Kee K, Angeles VT, Flores M, Nguyen HN, Reijo Pera RA. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature* 2009; 462:222–225.
21. Yokobayashi S, Okita K, Nakagawa M, Nakamura T, Yabuta Y, Yamamoto T, Saitou M. Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate. *Biol Reprod* 2017; 96:1154–1166.
22. Diaz Perez SV, Kim R, Li Z, Marquez VE, Patel S, Plath K, Clark AT. Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. *Hum Mol Genet* 2012; 21:751–764.
23. Patel S, Bonora G, Sahakyan A, Kim R, Chronis C, Langerman J, Fitz-Gibbon S, Rubbi L, Skelton RJP, Ardehali R, Pellegrini M, Lowry WE et al. Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep* 2017; 18:54–67.
24. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; 339:819–823.
25. Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M, Clark AT. 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* 2013; 15:113–122.
26. Li Z, Yu J, Hosohama L, Nee K, Gkountela S, Chaudhari S, Cass AA, Xiao X, Clark AT. The Sm protein methyltransferase PRMT5 is not required for primordial germ cell specification in mice. *EMBO J* 2015; 34:748–758.
27. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; 25:1105–1111.
28. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; 31:166–169.
29. Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. A unique gene regulatory network resets the human germline epigenome for development. *Cell* 2015; 161:1453–1467.
30. Tamura RN. Epithelial integrin alpha 6 beta 4: complete primary structure of alpha 6 and variant forms of beta 4. *J Cell Biol* 1990; 111:1593–1604.
31. Trzpis M, McLaughlin PMJ, de Leij LMFH, Harmsen MC. Epithelial cell adhesion molecule. *Am J Pathol* 2007; 171:386–395.
32. Gkountela S, Zhang KX, Shafiq TA, Liao W, Hargan-Calvopiña J, Chen P, Clark AT. DNA demethylation dynamics in the human prenatal germline. *Cell* 2015; 161:1425–1436.
33. Loveland KL, Schlatt S. Stem cell factor and c-kit in the mammalian testis: lessons originating from Mother Nature's gene knockouts. *J Endocrinol* 1997; 153:337–344.
34. Pesce M, Di Carlo A, De Felici M. The c-kit receptor is involved in the adhesion of mouse primordial germ cells to somatic cells in culture. *Mech Dev* 1997; 68:37–44.
35. Richardson BE, Lehmann R. Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat Rev Mol Cell Biol* 2010; 11:37–49.
36. Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, Yong J, Hu Y, Wang X, Wei Y, Wang W, Li R et al. The transcriptome and DNA methylation landscapes of human primordial germ cells. *Cell* 2015; 161:1437–1452.
37. Cruse G, Beaven MA, Music SC, Bradding P, Gillfillan AM, Metcalfe DD. The CD20 homologue MS4A4 directs trafficking of KIT toward clathrin-independent endocytosis pathways and thus regulates receptor signaling and recycling. *Mol Biol Cell* 2015; 26:1711–1727.
38. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated receptor for exploring gene sets in various biological contexts. *Nucleic Acids Res* 2005; 33:W741–W748.
39. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GENE ScT AnaLysis toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 2013; 41:W77–W83.
40. Funa NS, Schachter KA, Lerdrup M, Ekberg J, Hess K, Dietrich N, Honoré C, Hansen K, Semb H. β -Catenin regulates primitive streak induction through collaborative interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell* 2015; 16:639–652.
41. Inman GJ. SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol Pharmacol* 2002; 62:65–74.
42. Moon RT, Kohn AD, De Ferrari GV, Kaykas A. WNT and beta-catenin signalling: diseases and therapies. *Nat Rev Genet* 2004; 5:691–701.
43. Feng Q, Lu S, Klimanskaya I, Gomes I, Kim D, Chung Y, Honig GR, Kim K, Lanza R. Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells* 2010; 28:704–712.
44. Hu B-Y, Weick JP, Yu J, Ma L-X, Zhang X-Q, Thomson JA, Zhang S-C. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci USA* 2010; 107:4335–4340.
45. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggan K et al. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 2011; 144:439–452.
46. Butcher LM, Ito M, Brimpari M, Morris TJ, Soares FAC, Åhrlund-Richter L, Carey N, Vallier L, Ferguson-Smith AC, Beck S. Non-CG DNA methylation is a biomarker for assessing endodermal differentiation capacity in pluripotent stem cells. *Nat Commun* 2016; 7:10458.
47. Russ AP, Wattler S, Colledge WH, Aparicio SA, Carlton MB, Pearce JJ, Barton SC, Surani MA, Ryan K, Nehls MC, Wilson V, Evans MJ. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* 2000; 404:95–99.
48. Kurimoto K, Saitou M. Mechanism and reconstitution in vitro of germ cell development in mammals. *Cold Spring Harb Symp Quant Biol* 2015; 80:147–154.

Chapter 4

An extended culture system that supports human primordial germ cell-like survival and initiation of DNA methylation erasure

An Extended Culture System that Supports Human Primordial Germ Cell-like Cell Survival and Initiation of DNA Methylation Erasure

Joanna J. Gell,^{3,4,5,10,11} Wanlu Liu,⁶ Enrique Sosa,^{1,3} Alex Chialastri,^{7,8} Grace Hancock,^{1,2,3} Yu Tao,^{1,3} Sissy E. Wamaitha,^{1,3} Grace Bower,¹ Siddharth S. Dey,^{7,8,9} and Amander T. Clark^{1,2,3,5,*}

¹Department of Molecular Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA 90095, USA

²Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA 90095, USA

³Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA 90095, USA

⁴David Geffen School of Medicine, Department of Pediatrics, Division of Hematology-Oncology, Los Angeles, CA 90095, USA

⁵Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA 90095, USA

⁶Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Hangzhou 310058, P. R. China

⁷Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, CA 93106, USA

⁸Center for Bioengineering, University of California Santa Barbara, Santa Barbara, CA 93106, USA

⁹Neuroscience Research Institute, University of California Santa Barbara, Santa Barbara, CA 93106, USA

¹⁰Present address: Connecticut Children's Center for Cancer and Blood Disorders, Hartford, CT, USA

¹¹Present address: The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

*Correspondence: clarka@ucla.edu

<https://doi.org/10.1016/j.stemcr.2020.01.009>

SUMMARY

The development of an *in vitro* system in which human primordial germ cell-like cells (hPGCLCs) are generated from human pluripotent stem cells (hPSCs) has been invaluable to further our understanding of human primordial germ cell (hPGC) specification. However, the means to evaluate the next fundamental steps in germ cell development have not been well established. In this study we describe a two dimensional extended culture system that promotes proliferation of specified hPGCLCs, without reversion to a pluripotent state. We demonstrate that hPGCLCs in extended culture undergo partial epigenetic reprogramming, mirroring events described in hPGCs *in vivo*, including a genome-wide reduction in DNA methylation and maintenance of depleted H3K9me2. This extended culture system provides a new approach for expanding the number of hPGCLCs for downstream technologies, including transplantation, molecular screening, or possibly the differentiation of hPGCLCs into gametes by *in vitro* gametogenesis.

INTRODUCTION

Primordial germ cells (PGCs) are the first germline embryonic progenitors in all metazoans. Once specified, PGCs are fate restricted to become mature gametes in adults. In mammals, PGC specification is followed by multiple key events as PGCs migrate from their initial site outside the embryo, into the hindgut and the genital ridges. During this time PGCs proliferate and undergo dramatic epigenetic reprogramming, including the global loss of methylated cytosines from DNA, and the dynamic loss of histone H3 lysine 9 dimethylation (H3K9me2) and gain of histone H3 lysine 27 trimethylation (H3K27me3) in PGC chromatin (Gkountela et al., 2013; Guo et al., 2015; Seisenberger et al., 2012; Seki et al., 2005, 2007; Tang et al., 2015). Once the PGCs have settled in the embryonic gonad, they will differentiate into male or female germ cells. Mouse models have revealed that abnormalities in PGC specification, proliferation, epigenetic reprogramming, and differentiation can lead to germ cell tumors, infertility, or the transmission of disease alleles to the next generation. Therefore, understanding PGC development is critical to understanding mechanisms responsible for mammalian fertility and to facilitate our understanding of human reproductive health.

Experimental strategies to investigate PGC proliferation, epigenetic reprogramming and differentiation in mammals has been hampered by a lack of approaches to support the long-term self-renewal of mouse (m) and human (h) PGCs *ex vivo*. Using the mouse, approaches for short-term culture of mPGCs or mPGC-like cells (mPGCLCs) differentiated from pluripotent stem cells have been described (Farini et al., 2005; Oliveros-Etter et al., 2015; Ohta et al., 2017). However, these approaches have limitations because removal of mPGCs from their embryonic environment results in either cell death or reversion into a pluripotent self-renewing cell type called embryonic germ cells (EGCs) (Durcova-Hills et al., 2001; Leitch et al., 2013; Matsui et al., 1992; Resnick et al., 1992). The ability of mPGCs to revert into mEGCs is age-dependent and coincident with expression of the pluripotent transcription factors Nanog, Oct4, and Sox2 during the mPGC stage beginning at embryonic day 7.5 (E7.5) (Leitch et al., 2013).

Evaluating the cell and molecular mechanisms that regulate hPGC development is challenging due to limited access to human embryonic and fetal tissue. A small number of studies have cultured hPGCs *ex vivo*; however, under these conditions the hPGCs either die or revert into hEGC-like cells (Hua et al., 2009; Liu et al., 2004; Shambloott et al., 1998; Turnpenny et al., 2003). Unlike the mouse,





where mEGCs exhibit long-term self-renewal, the reversion of hPGCs into hEGC-like cells is extremely inefficient, and hEGC-like cells cannot be sustained in long-term self-renewing conditions (Tumpenny et al., 2006).

Given this, hPGC development is routinely modeled using the differentiation of human pluripotent stem cells (hPSCs) into hPGCLCs, with the majority of studies using three-dimensional (3D) disorganized aggregates (Irie et al., 2015; Sasaki et al., 2015; Sybirna et al., 2019). These methods create early hPGCLCs equivalent to week 3 post-fertilization of human embryo development. Recent studies using microfluidics and the generation of modeled 3D embryos from hPSCs also results in the differentiation of early hPGCLCs (Zheng et al., 2019). However, analyzing hPGCLC biology beyond specification is a challenge in this microfluidic system, as the 3D modeled embryos can only be maintained for 48–72 h before the system collapses (Zheng et al., 2019). In contrast, hPGCLCs in the 3D disorganized aggregates can be maintained for about 2 weeks in the aggregate, and during this time the hPGCLCs undergo limited epigenetic reprogramming (Irie et al., 2015; Sasaki et al., 2015; von Meyenn et al., 2016). Given the relatively undefined nature of the somatic cells in the 3D disorganized aggregates, and the relatively short time in which the modeled 3D embryo technology from stem cells can be sustained, it could be proposed that neither model has been optimized for evaluating the cell and molecular events that occur after hPGCLC specification. Given this, we tested the ability of hPGCLCs to proliferate in extended culture as a means to promote the expansion of the hPGCLC population *in vitro* for downstream applications.

RESULTS

hPGCLCs Maintain hPGC Identity when Cultured on STO Cells

In this study, hPGCLCs were differentiated in disorganized 3D aggregates from primed human embryonic stem cells (hESCs) through incipient mesoderm-like cells (iMeLCs) following the protocol first developed by Sasaki et al. (2015) with minor alterations, such as omission of stem cell factor (SCF) from the PGCLC media (Chen et al., 2017) (Figure 1A). hPGCLCs were isolated from the aggregates at day 4 (D4) of aggregate differentiation using fluorescence-activated cell sorting (FACS) for integrin alpha 6 (ITGA6) and epithelial cell adhesion molecule (EPCAM) (Chen et al., 2017; Sasaki et al., 2015). FACS-isolated hPGCLCs were maintained on a feeder layer of STO cells for up to 21 additional days (D4C21), which would correspond to a total of 26 days from the undifferentiated hESC state (Figure 1A). We first evaluated two types of culture media for maintaining hPGCLC identity on STOs.

Seven-factor (7F) medium, which contains a complex recipe of cytokines and chemicals that were previously shown to be necessary for mPGC proliferation and survival (Farini et al., 2005). 7F medium has also been shown to support hPGCLC survival on polyethylene terephthalate membranes for 4 days (Gell et al., 2018). The second medium, called FR10, supports mPGCLC proliferation for 7 days on m220 feeders (Ohta et al., 2017). Both media contain the cyclic adenosine monophosphate agonist forskolin, in addition to the cytokine SCF (Table 1).

hPGCLCs were isolated from four different hESC lines and sublines, UCLA1 (U1) (46, XX), UCLA2 (U2) (46, XY), a UCLA2 subline called UCLA2-GFP where GFP is driven from the ubiquitin promoter, and UCLA6 (U6) (46, XY). Using the UCLA2-GFP subline, we observed small clusters of hPGCLC by day 3 of extended culture (D4C3) in both 7F and FR10 media (Figure 1B, represents FR10). Notably, the hPGCLCs were loosely attached on top of the STOs, forming grape-like clusters rather than typical flat primed hPSC colonies.

Given this unique morphology, we next performed immunofluorescence (IF) staining of U1, U2, and U6 hPGCLCs at D4C10 in 7F and FR10 media to evaluate hPGC identity (Figures 1C, S1A, and S1B). Germline identity was evaluated by triple staining for PRDM1, TFAP2C, and SOX17, which discriminate hPGCs *in vivo* from somatic cells (Chen et al., 2017). Using this strategy, we discovered that the majority of hPGCLCs were triple-positive in both 7F and FR10, with a small fraction of SOX17/TFAP2C double-positive hPGCLCs, which were more apparent in 7F relative to FR10 (Figure 1D). SOX17/PRDM1 double-positive cells were never observed (Figure 1D). Given the ability to maintain a greater percent of triple-positive hPGCLCs we continued the remainder of the study with FR10 media. To determine whether extended culture could be prolonged for additional days, we cultured hPGCLCs from the U1 and U2 hESC line to 21 days (D4C21) in FR10 media (Figure 1E, represents U1) and show that SOX17/TFAP2C/PRDM1 triple-positive hPGCLCs can be sustained for at least 21 days.

Because FR10 medium supports survival of E9.5 mPGCs on m220 feeders, albeit with limited proliferation (Ohta et al., 2017), we next evaluated whether FR10 medium supports the survival of hPGCs isolated from human embryonic gonads. To achieve this, we isolated TNAP/cKIT double-positive hPGCs by FACS from male E53 embryonic testes, and cultured the resulting E53 hPGCs on STOs in FR10 medium for 10 days. In FR10 medium, the E53 hPGCs remained round and loosely attached to the STOs as individual cells. Germline identity was maintained in the E53 hPGCs at day 10 (E53D10), as evident by IF staining for VASA and TFAP2C, while being negative for SOX2 (Figure S1C). This result suggests that the culture of E53 hPGCs

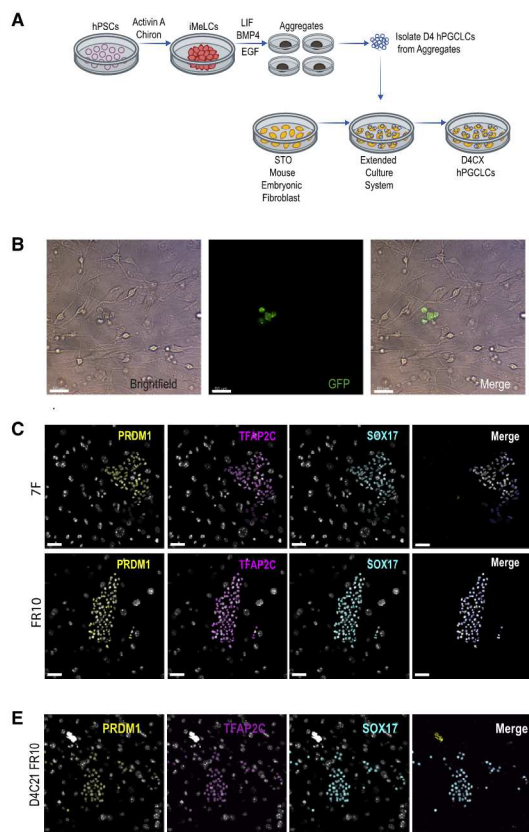


Figure 1. hPGCLCs Cultured on STOs Maintain Germline Identity

(A) Experimental scheme for extended culture of human primordial germ cell-like cells (hPGCLCs) on STOs. Day 4 (D4) hPGCLCs are maintained in culture for additional days (X) (D4CX). hPGCLCs in this study were cultured for a maximum of 21 days (D4C21).

(B) Bright-field (left), fluorescent microscopy (middle), and merged (right) images, illustrating UCLA2-GFP D4C3 hPGCLCs in culture on STOs in FR10 media.

(C) Immunofluorescent (IF) images of UCLA2 D4C10 hPGCLCs in 7-factor (7F) (top) and FR10 (bottom) media. Germline identity was evaluated using PRDM1 (yellow), TFAP2C (magenta), and SOX17 (cyan).

(D) Quantification of IF staining in UCLA2 D4C10 hPGCLCs for triple-positive SOX17/TFAP2C/PRDM1 (S/T/P) hPGCLCs and double-positive SOX17/PRDM1 (S/P) cells, or SOX17/TFAP2C (S/T) cells.

(E) IF images of UCLA1 D4C21 hPGCLCs in FR10 media. Germline identity denoted by triple-positive PRDM1 (yellow), TFAP2C (magenta), and SOX17 (cyan) cells. Scale bars, 50 μ m (B, C, and E).

in FR10 on STOs does not lead to reversion into hEGC-like cells during the first 10 days of extended culture.

Extended Culture Supports a Transcriptional Identity Similar to Early hPGCLCs

Given that hPGCLCs grew in clusters in extended culture, whereas hPGCs did not, we next sought to evaluate whether the hPGCLCs were acquiring markers of hEGCs. First, we performed IF at D4C10 for SOX2 (a EGC marker) and SOX17 (an hPGC marker). Undifferentiated hESCs were used as a positive control for SOX2. These results show that hPGCLCs at D4C10 are positive for SOX17 and do not express the hEGC marker SOX2 (Figure 2A, quantified in S2A).

Next, we developed a FACS strategy to isolate the cultured hPGCLCs from the STOs using fluorescent-labeled antibodies. This involved the use of an antibody that recog-

nized the surface molecule, cluster of differentiation 29 (CD29) to discriminate the mouse STOs together with an antibody that recognizes TRA-1-85, which discriminates human cells. Using this approach, we identified a population of CD29-positive mouse cells and TRA-1-85-positive human cells (Figure 2B). Using FACS to isolate the TRA-1-85-positive/CD29-negative cells, we performed RNA sequencing (RNA-seq) of UCLA1 and UCLA2 D4C10 putative hPGCLCs, and compared this with RNA-seq results from previously published data, including naive hESCs (Pastor et al., 2016), primed hESCs, iMeLCs, D4 hPGCLCs (Chen et al., 2017), and hPGCs isolated at various stages of germline cell development (Chen et al., 2017). Details on the RNA-seq libraries can be found in Table S1. Principle-component analysis (PCA) revealed that D4C10 hPGCLCs clustered together with D4 hPGCLCs in both principle component 1 (PC1) and PC2 (Figure 2C), and



Table 1. Key Components of 7F and FR10 Media. Shown in bold are media components common to both media types.

Media	7F	FR10
Growth factors and other additives	<ul style="list-style-type: none"> ● 50 ng/mL stem cell factor ● 10 ng/mL stromal cell-derived factor 1 ● 10 ng/mL fibroblast growth factor ● 25 ng/mL bone morphogenic protein 4 ● 10 µg/mL leukemia inhibitory factor ● 100 µg/mL N-acetylcysteine ● 5 µM forskolin 	<ul style="list-style-type: none"> ● 100 ng/mL stem cell factor ● 10 µM forskolin ● 10 µM rolipram

that germline identity of the D4C10 hPGCLCs is equivalent to early hPGCs but not late PGCs that have colonized the gonad (Figure 2D). Real-time PCR was used to confirm the RNA-seq results showing that D4C10 hPGCLCs expressed equivalent levels of *PRDM1*, *SOX17*, *TFAP2C*, and *NANOS3* to D4 hPGCLCs, and that *SOX2* mRNA was not detected (Figure S2B). These results suggest that the extended culture system maintains hPGCLC identity at a stage equivalent to early hPGCs, and does not promote reversion to self-renewing pluripotent primed or naive hES-like cells.

Although the overall transcriptional identity of hPGCLCs at D4C10 was similar to D4 hPGCLCs, we identified a small number of differentially expressed genes when performing pairwise comparisons with primed hESCs, D4 hPGCLCs and hPGCs (Figure 2E). Notably, this result shows that extended culture leads to considerably more upregulated genes in the D4C10 hPGCLCs relative to downregulated genes in each pairwise comparison (Figure 2E; Chart S1). Furthermore, gene ontology (GO) (Figure 2F) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (Figure S2C) of the 675 upregulated genes in D4C10 hPGCLCs relative to D4 hPGCLCs reveals biological terms, including extracellular matrix, endoplasmic reticulum lumen, and a variety of cell signaling pathways, indicating that the hPGCLCs in extended culture are responding to their new culture environment.

Extended Culture on STO Cells Supports hPGCLC Proliferation

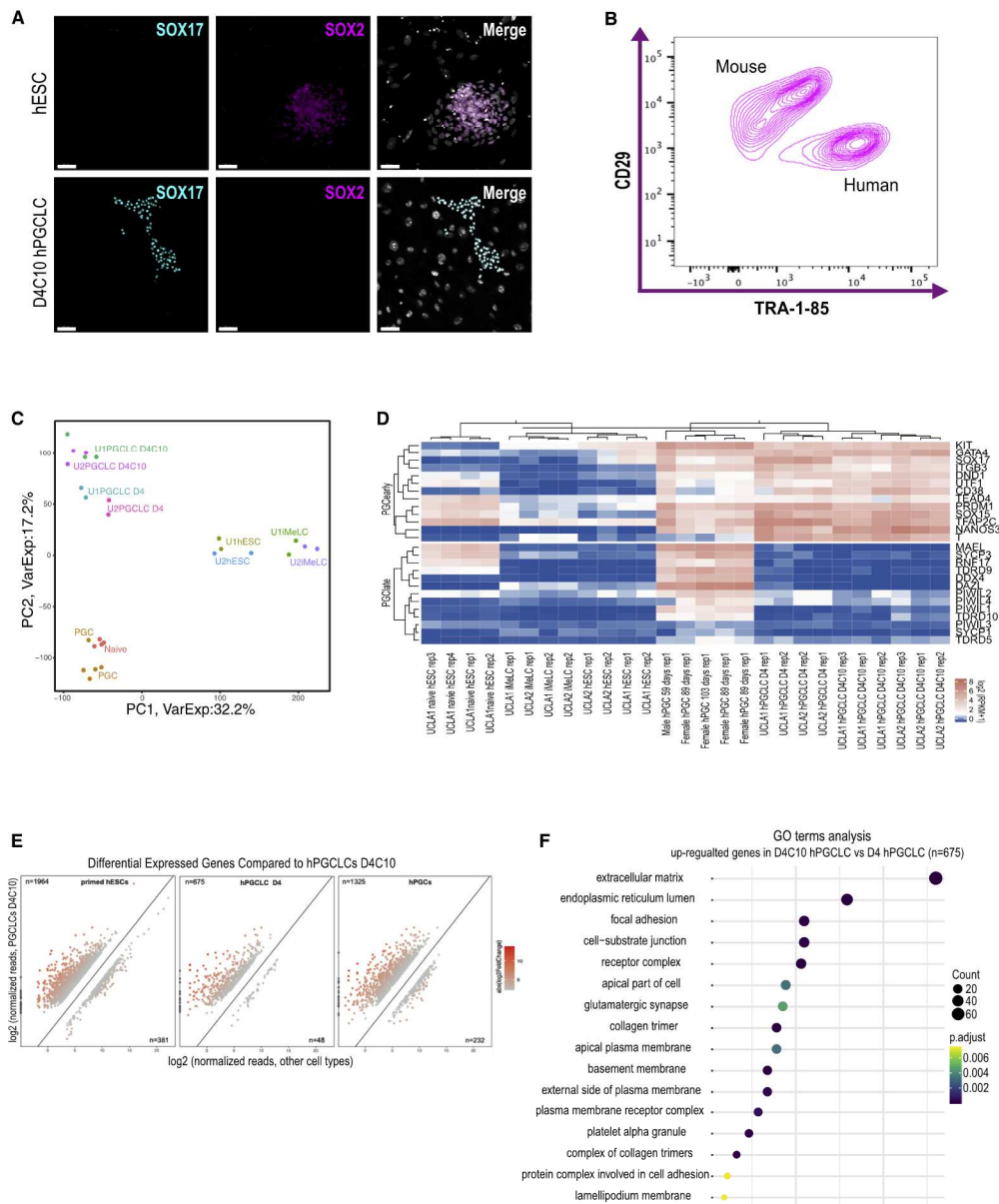
Previous studies using FR10 to culture mPGCLCs resulted in 20- to 50-fold expansion in mPGCLC numbers (Ohta et al., 2017). To quantify the total number of hPGCLCs at D4C10 and D4C21 we used FACS to isolate and count the number of hPGCLCs isolated from the STOs at each time point. We compared these values with the initial number

of D4 hPGCLCs plated into culture (Figure 3A). At D4C10 the UCLA1 and UCLA2 D4C10 hPGCLCs showed only modest capacity for expansion, increasing around 1.5- to 2-fold in cell number (Figure 3A). We do not think this modest increase is due to an increase in apoptosis as apoptotic genes are not differentially expressed (Figure S3A). However, by D4C21 the number of TRA-1-85 cells had increased by 25-fold (Figure 3A).

Using Ki67, we next quantified the percentage of hPGCLCs in cycle (Ki67+) at D4 in the aggregates and then in extended culture at D4C10 and D4C21. To identify D4 hPGCLCs in the aggregates, we quantified SOX17 and TFAP2C double positive hPGCLCs, which revealed that D4 hPGCLCs are mostly out of the cell cycle (Figure 3B, quantified in 3C). In contrast, in extended culture there was a statistically significant increase in the percentage of SOX17-positive hPGCLCs in cycle, with 60% of D4C10 and D4C21 hPGCLCs expressing Ki67 (Figure 3C). To evaluate progression through S phase, we performed a 5-ethynyl-2'-deoxyuridine staining on D4 aggregates and D4C10 and D4C21 of extended culture (Figure 3D, quantified in 3E). This result shows that, at D4, 25% of hPGCLCs are in S phase, which corresponds to the majority of cycling hPGCLCs in the aggregate. In the extended culture our results suggest that the percentage of cells in S phase remains the same at around 30%.

Extended Culture Maintains Partial Histone Reprogramming

Given that hPGCLCs are proliferating in extended culture, we next sought to evaluate epigenetic reprogramming. In the mouse embryo, H3K9me2 is one of the earliest histone modifications to be depleted from chromatin soon after mPGC specification (Kurimoto et al., 2015; Prokopuk et al., 2017; Seki et al., 2005). To evaluate H3K9me2, we compared H3K9me2 IF intensity in D4 aggregate hPGCLCs and at D4C10 and D4C21 of extended culture. Fluorescence intensity was normalized relative to the D4 aggregate somatic cells (Figure 4A, quantified in 4B). Consistent with previous reports (Sasaki et al., 2015), H3K9me2 was reduced in hPGCLCs in the aggregate compared with the somatic cells, and these reduced levels were maintained in extended culture. Because hPGCs are enriched in H3K27me3 (Gkountela et al., 2013; Guo et al., 2015; Tang et al., 2015), we analyzed H3K27me3 in hPGCLCs at D4, D4C10, and D4C21 relative to D4 somatic cells (Figure 4C, quantified in 4D). These results revealed an enrichment in H3K27me3 in D4 hPGCLCs relative to somatic cells. However, with extended culture to day D4C21, H3K27me3 levels were reduced to similar levels found in somatic cells of the aggregate. We also evaluated H3K9me2 and H3K27me3 in D4C10 hPGCLCs from the UCLA1 and UCLA6 hESC lines, which gave similar results (Figures S4A and S4B).



(legend on next page)



Extended Culture Supports Heterogeneous DNA Demethylation of hPGCLCs

Given that reduced H3K9me2 levels were maintained in hPGCLCs during extended culture, we next evaluated DNA methylation. In the mouse, loss of DNA methylation in mPGCs is hypothesized to be due to repression of UHRF1 protein and loss of replication-coupled DNA methylation maintenance (Kagiwada et al., 2013). Furthermore, UHRF1 protein is also repressed in hPGCs (Gkoutela et al., 2015; Tang et al., 2015). Using IF, we found that UHRF1 protein is not detectable in hPGCLCs at D4 as reported previously (Sasaki et al., 2015) and remains repressed during extended culture (Figure 5A).

Given the repression of UHRF1 protein in hPGCLCs, we next evaluated expression of the DNA methyltransferases (DNMT) (Figure S5A) and ten-eleven translocation 1-3 (TET1-3) genes (Figure S5B). We discovered that *DNMT1* and *DNMT3A* mRNA are expressed by hPGCs and D4C10 hPGCLCs, whereas *DNMT3B* and *DNMT3L* levels are reduced. This result suggests a reduction in *de novo* DNA methylation activity in the hPGCLCs during extended culture. In addition, our results show that the *TET* genes are expressed at similar levels in D4C10 hPGCLCs relative to the levels in hPGCs. These results suggest the potential for some loss of DNA methylation in the hPGCLCs during extended culture.

To quantify DNA methylation, we performed whole-genome bisulfite sequencing (WGBS) of UCLA2 hESCs, hPGCLCs at D4 and hPGCLCs cultured to D4C10. Averaging all CG methylation in each biological replicate revealed that undifferentiated hESCs had on average 80% CG DNA methylation, with comparable levels in D4 hPGCLCs (Figure 5B). In contrast the average CG DNA methylation at D4C10 was reduced to around 60% (Figure 5B). Consistent with the RNA-seq data showing repression of *DNMT3L* and *DNMT3B* in hPGCLCs with extended

culture, we also found that non-CG methylation was reduced in hPGCLCs relative to undifferentiated hESCs (Figures S5C–S5E).

In previous studies analyzing imprint demethylation in hPGCLCs differentiated from naive hESCs could not be performed because naive hESCs have eroded imprints (Pastor et al., 2016; von Meyenn et al., 2016). In the primed state, we identified 31 germline imprinted regions with an average CG DNA methylation of ~50% (Figure 5C). In D4 and D4C10 hPGCLCs, we discovered that the average CG methylation over these primary imprinting control regions was equivalent (Figure 5C). Therefore, although in bulk WGBS we can detect a modest reduction in global DNA demethylation, imprinting control regions remain methylated.

Given that almost 50% of the genome is composed of transposons we next evaluated average DNA methylation at long interspersed nuclear elements (LINEs), short interspersed nuclear elements, and long terminal repeats (Figures 5D and S5F–S5I). The DNA methylation levels of major transposon classes were equivalent to the genome average for all samples. An exception to this is the “escaper” LINE1 human-specific (L1HS) retrotransposon family, which is more resistant to DNA demethylation in hPGCs (Gkoutela et al., 2015). Our results show that in D4C10, L1HS retrotransposons have similar DNA methylation levels to D4 hPGCLCs, whereas an evolutionary older descendant, L1PA8, has significantly reduced DNA methylation relative to D4 hPGCLCs (Figures S5H and S5I). Protein coding genes also exhibited DNA demethylation in D4C10 hPGCLCs at the promoter and along the gene body compared with D4 hPGCLCs (Figure 5E). However, the transcription start site (TSS) was similarly demethylated in all samples. Together, these data show that, in the extended culture, hPGCLCs undergo modest genome-wide DNA demethylation. However, imprinted genes and L1HS remain methylated.

Figure 2. hPGCLCs in Extended Culture Maintain a Transcriptome Similar to Specified hPGCLCs

(A) IF images of primed UCLA2 hESCs (top) and D4C10 hPGCLCs in FR10 media (bottom). Pluripotency/EGC identity was evaluated using SOX2 (magenta). SOX17 (cyan) identifies the hPGCLCs. Scale bars, 50 μ m.

(B) FACS plot of UCLA2 D4C10 hPGCLCs on ST0s, CD29-positive mouse cells and TRA-1-85-positive human cells form two distinct populations.

(C) Principle-component analysis (PCA) of the transcriptomes of UCLA1 (U1) naive hESCs in 5i/L/FA (N = 4); UCLA1 and UCLA2: iMeLcs (N = 2), U1 and U2 primed hESCs (N = 2), U1 and U2 D4 hPGCLCs (N = 3), U1 and U2 D4C10 hPGCLCs (N = 3); and gonadal hPGCs (N = 5). Gene expression analysis includes RNA-seq data from Pastor et al. (2016) (5i/L/FA naive hESCs) and Chen et al. (2017) (primed hESCs, iMeLcs, D4 hPGCLCs, and hPGCs). N = independent replicates.

(D) Heatmap of gene expression levels of representative genes in UCLA1 naive hESCs in 5i/L/FA; UCLA1 and UCLA2: iMeLcs, primed hESCs and D4 hPGCLCs and D4C10 hPGCLCs; gonadal hPGCs. Genes evaluated are grouped as diagnostic for early hPGCs (i.e., SOX17, PRDM1, and TFAP2C) and late hPGC (i.e., DAZL, DDX4, and SYCP3). Rep, independent replicates.

(E) Differentially expressed genes (DEGs) in D4C10 hPGCLCs compared to primed hESCs (left), D4 hPGCLCs (center) and gonadal hPGCs (right). Using DEGs with fold change ≥ 4 , false discovery rate < 0.05 , and average RPKM in either cell types must be > 1 .

(F) Dot plot depicting gene ontology (GO) terms identified for the N = 675 upregulated genes in D4C10 hPGCLCs versus D4 hPGCLCs.

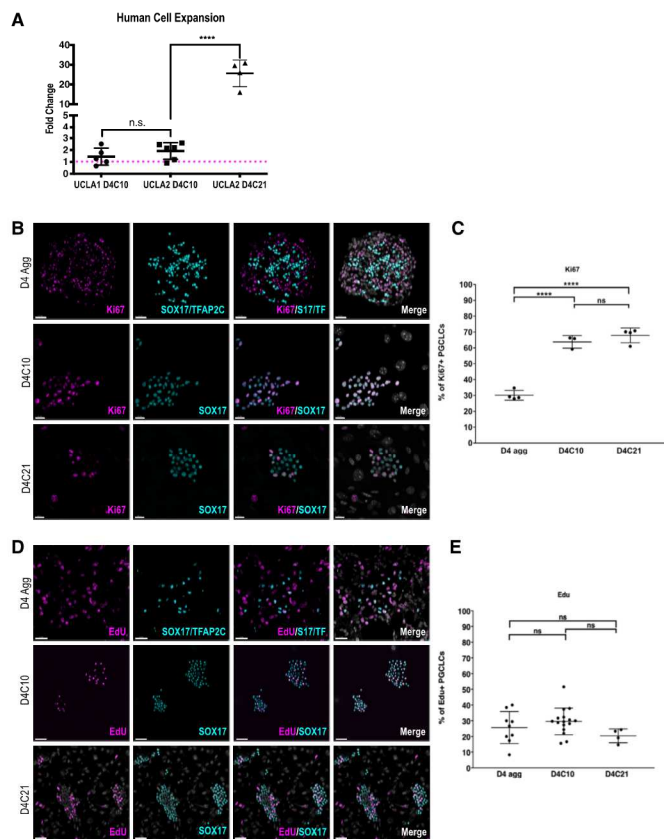


Figure 3. hPGCLCs in Extended Culture Self-Renew and Undergo Expansion

(A) Dot plot of the fold change in FACS isolated D4C10 and D4C21 hPGCLCs, compared with the starting D4 hPGCLCs. UCLA1 D4C10 (n = 5); UCLA2 D4C10 (n = 6) and D4C21 (n = 4). N = technical replicates. UCLA1 D4C10 and UCLA2 D4C10 represent 3 independent experiments, UCLA2 D4C21 represent 2 independent experiments. Magenta dotted line represents a fold change of 1. N.S., not significant; ****p ≤ 0.0001. Error bars = mean SD.

(B) IF images of Ki67 (magenta) in UCLA2 D4 aggregate hPGCLCs, marked by SOX17/TFAP2C (cyan) (top panel), D4C10 hPGCLCs, marked by SOX17 (cyan) (middle panel), and D4C21 hPGCLCs, marked by SOX17 (cyan) (bottom panel). S17/TF = SOX17/TFAP2C. Scale bars, 50 μm (top), 20 μm (middle), and 20 μm (bottom).

(C) Quantification of Ki67-positive hPGCLCs in UCLA2 D4 aggregates (D4 Agg), D4C10, and D4C21 hPGCLCs in FR10 media. ****p ≤ 0.0001. Error bars = mean SD for D4 Agg (N = 4), D4C10 (N = 3), and D4C21 (N = 3), N, the number of independent biological replicates.

(D) IF images of 5-ethynyl-2'-deoxyuridine (EdU) (magenta) in UCLA2 D4 aggregate hPGCLCs, marked by SOX17/TFAP2C (cyan) (top panel), D4C10 hPGCLCs, marked by SOX17 (cyan) (middle panel), and D4C21 hPGCLCs, marked by SOX17 (cyan) (bottom panel). S17/TF = SOX17/TFAP2C. Scale bars, 30 μm (top), 50 μm (middle), and 50 μm (bottom).

(E) Quantification of EdU-positive hPGCLCs in UCLA2 D4 aggregates (D4 Agg), D4C10, and D4C21 hPGCLCs in FR10 media. N.S., not significant. Error bars = mean SD of 9 aggregates (D4 Agg), 16 colonies (D4C10), and 4 colonies (D4C21) from 3 independent biological replicates.

Because UHRF1 protein is not detectable in hPGCLCs and around 30% of hPGCLCs in extended culture are in S phase, we evaluated DNA methylation in single cells, with the hypothesis that hPGCLCs at D4C10 are heterogeneously demethylating, meaning that some cells are initiating DNA demethylation while other cells are not. Utilizing a strand-specific, enzymatic-based method of sequencing, we compared the 5mC content of the plus strand relative to the whole chromosome in n = 84 single hPGCLCs at D4 and n = 68 single hPGCLCs at D4C10. This calculation is denoted as strand bias (*f*) ($f = 5 \text{ mC on + strand} / \text{total } 5 \text{ mC on chromosome}$). Calculation of strand bias in individual cells allowed for evaluation of replication-coupled DNA (demethylation with 5mC main-

tenance represented by a strand bias value of 0.5. A failure to maintain 5 mC during replication would cause an increase in strand bias variance at individual chromosomes. This experiment revealed that a small number of single cells at both D4 and D4C10 exhibit strand bias variance deviating from 0.5, with a trend for more D4C10 hPGCLCs exhibiting strand bias variance, and therefore a failure to maintain DNA methylation during DNA replication (Figures 5F and 5G). Using a tSNE plot, the single-cell data are displayed as D4 hPGCLCs (dots) and D4C10 hPGCLCs (triangles) (Figure 5F). Cluster 1 represents cells with low strand bias variance with 5mC maintained on both strands (Figures 5F and 5Sj). Cluster 2 represents cells with higher strand bias variance, such that the cells with the highest

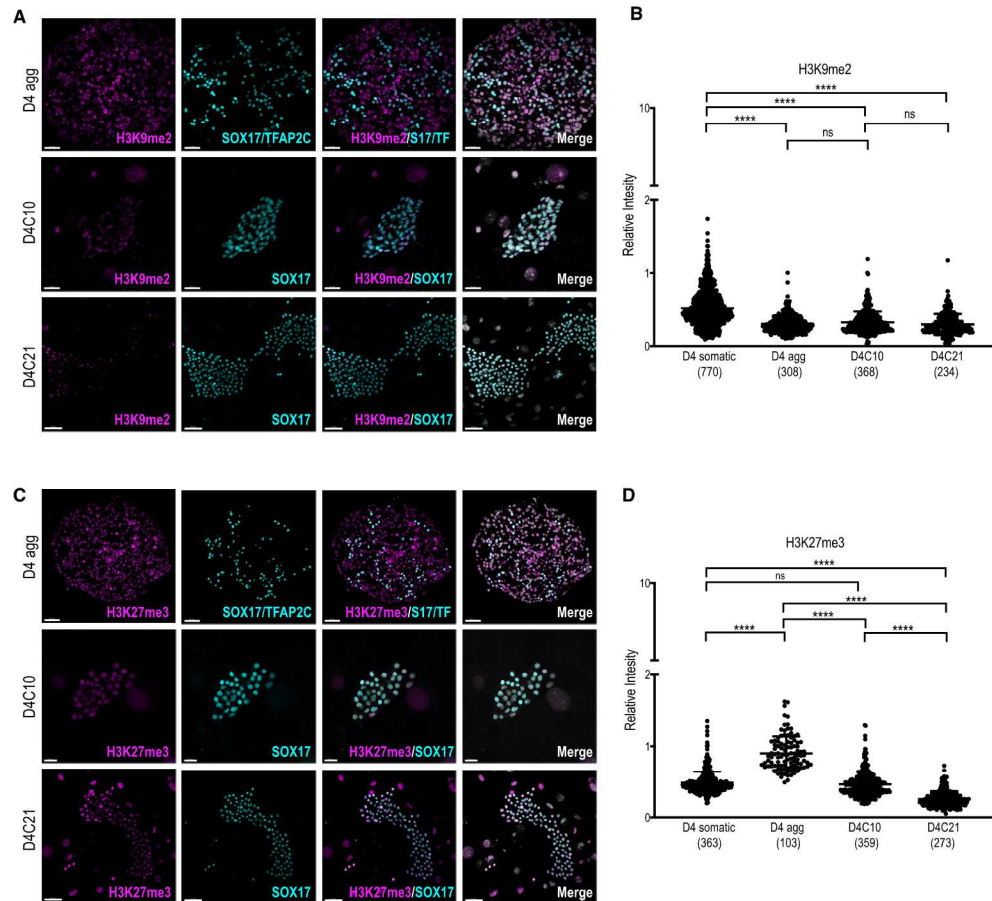


Figure 4. Partial Chromatin Remodeling Is Maintained in Extended Culture hPGCLCs

(A) IF images of H3K9me2 in aggregates containing UCLA2 D4 hPGCLCs, marked by SOX17/TFAP2C (cyan) (top panel), D4C10 hPGCLCs, marked by SOX17 (cyan) (middle panel), and D4C21 hPGCLCs, marked by SOX17 (cyan) (bottom panel). S17/TF = SOX17/TFAP2C. Scale bars, 40 μ m (top), 30 μ m (middle), and 50 μ m (bottom).

(B) Quantification of H3K9me2 in UCLA2 aggregates containing D4 hPGCLCs (D4 agg), D4C10 hPGCLCs, and D4C21 hPGCLCs in FR10 medium, as compared with H3K9me2 in D4 agg somatic cells, relative to DAPI fluorescence intensity. N.S., not significant; **** $p < 0.0001$. Error bars = mean SD of three independent biological replicates. Numbers in parentheses are equal to the total number of cells analyzed.

(C) IF images of H3K27me3 in UCLA2 aggregates containing D4 hPGCLCs, marked by SOX17/TFAP2C (cyan) (top panel), D4C10 hPGCLCs, marked by SOX17 (cyan) (middle panel), and D4C21 hPGCLCs, marked by SOX17 (cyan) (bottom panel). S17/TF = SOX17/TFAP2C. Scale bars, 50 μ m (top), 20 μ m (middle), and 50 μ m (bottom).

(D) Quantification of H3K27me3 levels in D4 hPGCLCs in the aggregate (D4 agg), U2 D4C10 hPGCLCs, and U2 D4C21 hPGCLCs in FR10 medium, as compared with H3K27me3 levels of D4 agg somatic cells, relative to DAPI fluorescence intensity. **** $p < 0.0001$. Error bars = mean SD of three independent biological replicates. Numbers in parentheses are equal to the total number of cells analyzed.

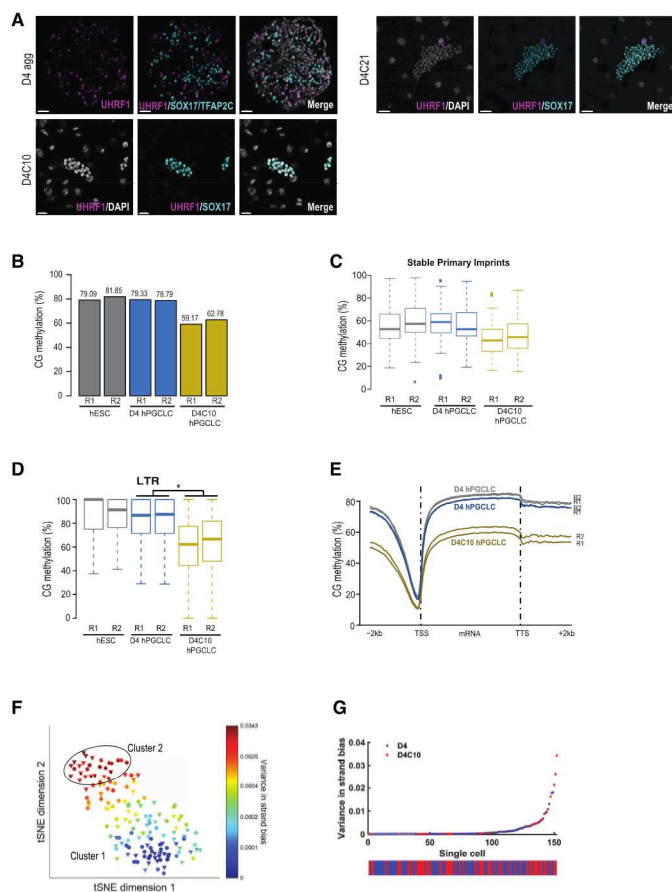


Figure 5. hPGCLCs in Extended Culture Undergo Partial Genome-Wide Demethylation

(A) IF images of UHRF1 expression in UCL A2 (U2) D4 aggregate hPGCLCs, marked by SOX17/TFAP2C (cyan) (top left panel), U2 D4C10 hPGCLCs, marked by SOX17 (cyan) (bottom left panel), and U2 D4C21 hPGCLCs, marked by SOX17 (cyan) (top right panel). Scale bars, 50 μ m.

(B) Bar graph of average percent CG methylation in UCL A2: hESCs (gray), D4 hPGCLCs (blue), and D4C10 hPGCLCs (yellow).

(C) Boxplot of CG methylation percentages in U2 hESCs (gray), U2 D4 hPGCLCs (blue), and U2 D4C10 hPGCLCs (yellow) over primary imprints (N = 31).

(D) Boxplot of CG methylation percentages over long terminal repeats (LTRs). U2 hESCs (gray), U2 D4 hPGCLCs (blue), and U2 D4C10 hPGCLCs (yellow). * $p < 2.2 \times 10^{-16}$.

(E) Metaplot of percent CG methylation over protein coding genes and flanking 2-kb regions in U2 hESCs (gray), U2 D4 hPGCLCs (blue), and U2 D4C10 hPGCLCs (yellow). TSS, transcription start site; TTS, transcription termination site.

Replicates for (B–E) are independent replicates. R1, replicate 1; R2, replicate 2.

(F) t-SNE plot based of the strand bias of single-cell U2 D4 aggregate hPGCLCs (dots) and U2 D4C10 hPGCLCs cultured in FR10 (triangles). Cluster 1, low variance; cluster 2, high variance. Black circle indicate cells with the highest strand bias variance.

(G) Dot plot of each individual U2 D4 hPGCLC (blue) and U2 D4C10 hPGCLC (red), ordered by variance in strand bias.

strand bias variance represent those cells that have lost 5mC on one strand while retaining 5 mC on the other (black circle) (Figures 5F and S5K). A scatterplot illustrating each individual D4 (blue) and D4C10 (red) hPGCLC further illustrates that increased variance in strand bias occurs in the D4C10 hPGCLCs (Figure 5G). Taken together, culturing D4 hPGCLCs in extended culture for 10 days leads to heterogeneous replication-coupled DNA demethylation.

DISCUSSION

In this paper we sought to develop an *in vitro* culture system with the capability to maintain hPGCLC identity

while promoting proliferation. Previous attempts to culture hPGCs from fetal gonads have failed to establish cell lines that maintain germline identity, and instead resulted in the formation of hEGC-like cells that cannot be maintained in culture (He et al., 2007; Hua et al., 2009; Liu et al., 2004; Shamlott et al., 1998; Turnpenny et al., 2003). Promoting hPGCLCs to proliferate while maintaining germline identity provides the opportunity for future applications, such as molecular screening, transplantation, or molecular analyses, where larger numbers of cells are required. Previous reports have suggested that D8 hPGCLCs are capable of some epigenetic reprogramming (Sasaki et al., 2015). Indeed, our results confirm that the loss of H3K9me2 is initiated by D4 while the hPGCLCs are still in the aggregate. However, a distinct



advantage of using extended culture is the maintenance of hPGCLCs for at least 3 weeks. In particular, extended culture could be used downstream of 3D modeled embryos from stem cells cultured microfluidic devices which only last 48–72 h (Zheng et al., 2019).

RNA-seq of D4C10 hPGCLCs revealed a transcriptome similar to D4 hPGCLCs, verifying the maintenance of germline identity rather than reversion to pluripotency as expected for EGCs. Furthermore, lack of SOX2 protein in the cultured hPGCLCs serves as an alternate approach for showing that extended culture does not promote the generation of hEG-like cells. One of the clear conclusions from the RNA-seq analysis is that the D4C10 hPGCLCs are not progressing into late stage hPGCs equivalent to hPGCs isolated from the fetal gonad. *In vivo*, the progression from early to late hPGCs in primates occurs around week 4–5 of embryo development, and is preceded by genome-wide DNA demethylation (Sasaki et al., 2016). It is anticipated that until genome-wide DNA demethylation is achieved, the early to late hPGCLC transition will not occur. Thus, an experimental approach for promoting additional DNA methylation reprogramming is required for germ cell development to progress in an orderly manner.

In this study, the hPGCLCs in extended culture had reduced H3K9me2, even after 21 days, combined with a partial loss of DNA methylation. The bulk levels of DNA methylation in our study were consistent with the work of von Meyenn et al. (2016) in which differentiation of hPGCLCs from naive hESCs through a re-primed intermediate showed a modest loss of DNA methylation by around 20%. In our study, we were able to extend this work by evaluating DNA methylation at germline imprints, which we found were protected from genome-wide DNA demethylation at D4C10. Consistent with the mouse data, this result suggests that the removal of DNA methylation from imprinted genes may be different from the mechanism that drives the initial stages of DNA demethylation in hPGCs (Hackett et al., 2013; Hargan-Calvopina et al., 2016; Hill et al., 2018; Vincent et al., 2013). Regarding the partial genome-wide loss of DNA methylation, our data suggest that this is due to heterogeneous replication-coupled DNA demethylation in some cells but not others. Future studies are necessary to capture the full DNA demethylation expected of hPGCs *in vivo*.

In summary, we report an extended culture system for hPGCLCs that retains germline identity while also recapitulating the earliest stages of epigenetic reprogramming. This 2D culture platform is an important new tool as it allows for new innovations in understanding imprint erasure in the human germline and efficient establishment of the late-stage gonocyte program, which has so

far eluded stem cell biologists interested in germline cell differentiation. Achieving these two events will be critical before initiating sex-specific differentiation toward eggs and sperm.

EXPERIMENTAL PROCEDURES

hPGCLCs or hPGCs Cultured on STOs

Human ESCs were cultured as previously described (Gell et al., 2018). Human fetal tissue was acquired and staged as previously described (Gkoutela et al., 2013, 2015). Sorted hPGCLCs or hPGCs were cultured in either 7F or FR10 medium as indicated. 7F media contents were prepared as described previously, with key components given in Table 1 (Gell et al., 2018; Oliveros-Etter et al., 2015). FR10 medium (Ohta et al., 2017) contains 10% KSR, 2.5% FBS (Thermo Fisher Scientific, SH3007003), 1 × NEAA (Gibco, 11140-050), 1 mM sodium pyruvate (Gibco, 11360070), 2 mM L-glutamine (Gibco, 25030081), 0.1 mM 2-mercaptoethanol (Gibco, 21985-023), 1 × penicillin streptomycin (Gibco, 15140-122), 100 ng/mL SCF (PeproTech, 250-03), 10 μM forskolin (Sigma, F6886), 10 μM rolipram (Sigma, R6520), and 50 ng/mL primocin in Glasgow's MEM (Gibco, 11710-035). hPGCLCs were plated onto 200,000 cells/cm² of STOs. hPGCLCs were cultured on either 4-well glass chamber slides (Falcon, 08-774-209), 12- or 24-well plates. hPGCLCs were plated at a density of 1,000–3,000 cells per well for chamber slides and 24-well plates, and 10,000 cells per well for 12-well plates. Media were changed daily. For D4C21 hPGCLCs in chamber slides, STOs were replated at a density of 100,000 cells/cm² on day 10 of culture. Chamber slides were used for IF. Twelve- and 24-well plates were used for downstream experiments. All hESC experiments were reviewed and approved by the UCLA Embryonic Stem Cell Research Oversight Committee. All human fetal tissue research was reviewed and approved by the University of Washington IRB, no identifiers or codes accompanied the fetal tissue to UCLA. Therefore, the UCLA IRB determined that the fetal tissue research experiments at UCLA were not subject to additional human subjects review.

Quantification of hPGCLC Expansion in Extended Culture

D4 hPGCLCs were plated at a density of 1,000 cells in a 24-well plate containing STOs, this is denoted as culture day 0 (D4C0). For D4C21 cell counts, hPGCLCs were split on D4C10, using 0.05% trypsin for 5 min in 37°C. Cells were collected and centrifugation at 1.5k rpm for 5 min. Cells were resuspended in FR10 medium and passed through a 100-μm strainer. Cells were split 1 well to 2 wells. The total number of human cells in culture at D4C10 and D4C21 was quantified by collecting all TRA-1-85-positive cells by FACS (see further details in Supplemental Experimental Procedures). Fold change was calculated. Prism GraphPad was used to generate graphs and statistical analysis, with an unpaired t test.

Immunofluorescence Quantification

Image analyses and quantifications were performed with the image analysis software Imaris 9.3.1 (Bitplane). Fluorescence intensity values for epigenetic markers (H3K9me2 and H3K27me3) were



extracted using the surface-rendering feature to build surfaces for hPGCLCs and somatic cells. D4 hPGCLC were quantified by building surfaces for all TFAP2C/SOX17 double-positive nuclei, implementing the co-localization feature to make an hPGCLC co-localization channel, which excludes all single positive or negative for SOX17 and TFAP2C. For the somatic cell population, surfaces were built for all DAPI-positive nuclei that were not included in the hPGCLC group. To quantify hPGCLCs in extended culture, surfaces were built for all SOX17 cells that overlapped with DAPI signals. Mean fluorescence intensity (MFI) values were collected using the mask channel option on H3K9me2, H3K27me3, and DAPI channels. Relative intensity values for individual nuclei were calculated by dividing the MFI for the respective epigenetic mark by the MFI for DAPI. To quantify the proportion of Edu-positive and Ki67-positive cells, a similar procedure building surfaces was used to quantify hPGCLCs and somatic cells. The proportion of Edu-positive or Ki-67-positive hPGCLCs was obtained by dividing the number of hPGCLCs positive for either mark by the total number of DAPI-positive nuclei. Quantification of each group was based on confocal images of at least three individual aggregates or clusters of hPGCLCs for each time point. All IF experiments were performed as three independent replicates. Prism GraphPad was used to generate graphs and statistical analysis. Ordinary one-way ANOVA was used to compare the means between groups for statistical significance. To determine which groups differed from each other a Tukey's multiple comparison *post hoc* test was run.

Library Preparation and Sequencing

RNA was extracted using the RNeasy Micro Kit (QIAGEN) and quantified using a NanoDrop ND-1000 (NanoDrop). RNA sequencing libraries were prepared using the Nugen RNA-seq System V2 with 5–100 ng starting material. DNA was extracted using the Quick gDNA MiniPrep Kit (Zymo) and quantified using the Qubit dsDNA High-Sensitivity Kit (Life Technologies). Bisulfite-sequencing libraries were prepared using the Ovation Ultralow Methyl-Seq Library System (NuGEN). Unmethylated Lambda phage DNA (NEB) was spiked in at 0.25% input DNA quantity to determine conversion efficiency, which was 99.3%–99.5% for all libraries. Libraries were sequenced on Illumina HiSeq instruments. Details on the RNA-seq and WGBS libraries can be found in [Tables S1 and S2](#).

RNA-Seq Analysis

Analysis of Individual Gene Expression

RNA-seq data were analyzed as described previously ([Chen et al., 2017](#)). In general, raw paired end sequencing reads were first mapped to hg19 genome with TopHat ([Trapnell et al., 2009](#)) allowing up to two mismatches and one maximal multi-hit (-g 1). Read counts for each individual gene were calculated by HTseq ([Anders et al., 2015](#)) with default parameters. Expression levels of individual genes were calculated as RPKM (reads per kilobase of exons per million aligned reads) in R. For published datasets (GSE76970 and GSE93126) ([Chen et al., 2017](#); [Pastor et al., 2016](#)), processed data of the raw read counts for each gene were utilized, while the downstream analysis was processed in the same way as RNA-seq data generated from this study.

Heatmap

For plotting the heatmap of RNA-seq data in this study log₂ (RPKM+1) values for selected genes were used and plotted in R using the ComplexHeatmap package ([Gu et al., 2016](#)).

PCA

RPKM values for each gene across different samples were used as input to calculate the principal component. PCA analysis (prcomp function in R) was performed on all genes across different samples. PCA plots were then plotted with ggplot2 package in R (<http://ggplot2.org>).

Differential Expressed Genes

To define differential expressed genes, the DESeq package was utilized in R ([Anders and Huber, 2010](#)). Genes with average fold change > 4, adjusted p value < 0.05, and mean RPKM > 1 across replicates in either cell types for the comparison, were defined as differential expressed genes.

GO Term and KEGG Pathway Enrichment Analysis

To identify enriched GO term and KEGG pathway, analyses were performed using R package clusterProfiler with differential expressed genes called before ([Yu et al., 2012](#)).

WGBS Analysis

Raw reads of WGBS data were first aligned to hg19 genome using BSMAP ([Xi and Li, 2009](#)) by allowing up to two mismatches (-v2), maximal one equal best hit to count (-w 1) and mapping to both strands (-n 1). When calculating methylation level over each cytosine, only reads uniquely mapped were kept and PCR duplicates were removed. Methylation levels at CG sites were then calculated by #C/(#C+#T).

Metaplot of WGBS Data

For the metaplot of CG methylation over protein coding genes, all gene coordinates were utilized with 2 kb upstream TSS and 2 kb downstream of TSS. Then the 2-kb flanked genic regions were divided into 100-bp bins, and methylation level were calculated within each bin. Metaplots were then plotted in R with customized scripts.

Analysis on Transposons

To calculate CG methylation over transposons, transposons type and coordinates were downloaded from RepeatMasker for hg19 (<http://repeatmasker.org/>). The Wilcoxon rank-sum test was used for statistical analysis.

Analysis on Imprints

Coordinates for stable primary imprints and transient imprints were obtained from previously published data ([Okoe et al., 2014](#); [Pastor et al., 2016](#); [Smith et al., 2014](#)).

Whole-Genome Single-Cell 5mC Sequencing

(scMspJI-Seq)

Samples were prepared as described in scMspJI-seq with minor modifications ([Sen et al., 2019](#)). TRA-1-85-positive single hPGCLCs were sorted into each reaction well of a 384-well plate containing 4 μ L of Vapor-Lock (QIAGEN) and 200 nL of lysis buffer (0.0875% IGEPAL CA-630) and were stored at -80°C until use. Cells were prepared as described by [Sen et al. \(2019\)](#) with some minor adjustment in reagent volumes. The ds-adaptor sequences used are described in scMspJI-seq ([Sen et al., 2019](#)). Excluding the Vapor-Lock, all reagent



dispenses were performed using the Nanodrop II liquid handling robot (BioNex Solutions). Each plate was pooled into 4 libraries containing 96 uniquely barcoded cells. For each library, DNA clean-up was then performed with 1× Agencourt Ampure XP beads and eluted in 30 μL of nuclease-free water. The samples were vacuum centrifuged to a volume of 6.4 μL, and 9.6 μL of *in vitro* transcription mix was added (1.6 μL of each ribonucleotide, 1.6 μL of T7 buffer, 1.6 μL T7 enzyme mix [MEGAscript T7 Transcription Kit, Ambion]) and incubated at 37°C for 13 h. Library preparation steps were performed as described previously (Mooijman et al., 2016; Rooijers et al., 2019). The identification of 5 mC sites in the human genome from the sequencing data is described in scMspJI-seq (Sen et al., 2019).

Analysis on scMspJI-Seq

MspJI recognizes 5 mC sites and cuts gDNA 16 bp downstream of the site leaving a 4-bp 5′ overhang, allowing both position-specific and strand-specific detection of 5 mC. The strand-specific distribution of 5 mC on a chromosome is quantitatively described by a metric called strand bias, as described previously (Sen et al., 2019; Mooijman et al., 2016). Ninety-six U2 D4 hPGCLCs and 94 U2 D4C10 hPGCLCs cells were sequenced, along with 2 negative control empty reaction wells. Poorly sequenced cells containing less than 10,000 unique 5 mCG sites were removed (median detection 60,597 unique 5 mCG sites per cell), leaving 84 U2 D4 hPGCLCs and 68 U2 D4C10 hPGCLCs passing quality control. A 5 mCG strand bias was calculated for each chromosome in all cells. The variance in the strand bias of chromosomes within each cell was used as a measure of DNA methylation maintenance. Mean silhouette scores were used to identify two populations of cells from k-means clustering on the 5 mCG strand bias, a population with low variance in the strand bias distribution, and another population with high variance in the strand bias distribution (Sen et al., 2019).

ACCESSION NUMBERS

All data mentioned in the paper are deposited in the NCBI GEO. RNA-seq and WGBS under accession number GSE139115, scMspJI-seq under accession number GSE138982.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.stemcr.2020.01.009>.

AUTHOR CONTRIBUTIONS

J.J.G. conceived and performed the experiments and wrote the manuscript. W.L. analyzed the sequencing data. Y.T. performed preparation of the WGBS libraries. E.S., G.H., and G.B. performed the IF experiments. E.S. quantified the IF experiments. S.E.W. quantified the hPGCLCs in culture. A.C. and S.S.D. performed and analyzed the single-cell 5 mC sequencing. A.T.C. conceived the experiments, wrote the manuscript, and maintained the Embryonic Stem Cell Research Oversight for hESCs.

ACKNOWLEDGMENTS

This project was funded with supported from the Pablove Foundation 20183715 (to J.J.G.), R01HD079546 (to A.T.C.), and supplement to R01 R01HD058047 (to A.T.C.), and UC Cancer Research Coordinating Committee (CTN-19-585462) grant to S.S.D. G.H. acknowledges the support of the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Training Program. The authors thank the following UCLA Broad Stem Cell core facilities: The Imaging core, Flow cytometry core, Genomics core, and the Stem Cell Banking core. The authors acknowledge the use of the Biological Nanostructures Laboratory within the California NanoSystems Institute for Illumina sequencing, supported by the University of California, Santa Barbara and the University of California, Office of the President. Human fetal tissue procurement in this research was supported by 5R24HD000836. Requests for fetal tissues can be made to Ian Glass at the University of Washington Birth Defects Laboratory.

Received: April 22, 2019

Revised: January 8, 2020

Accepted: January 14, 2020

Published: February 13, 2020

REFERENCES

- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* *11*, R106.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Chen, D., Liu, W., Lukianchikov, A., Hancock, G.V., Zimmerman, J., Lowe, M.G., Kim, R., Galic, Z., Irie, N., Surani, M.A., et al. (2017). Germline competency of human embryonic stem cells depends on eomesodermin. *Biol. Reprod.* *97*, 850–861.
- Durcova-Hills, G., Ainscough, J., and McLaren, A. (2001). Pluripotential stem cells derived from migrating primordial germ cells. *Differentiation* *68*, 220–226.
- Farini, D., Scalfaferrri, M.L., Iona, S., La Sala, G., and De Felici, M. (2005). Growth factors sustain primordial germ cell survival, proliferation and entering into meiosis in the absence of somatic cells. *Dev. Biol.* *285*, 49–56.
- Gell, J.J., Zhao, J., Chen, D., Hunt, T.J., and Clark, A.T. (2018). PRDM14 is expressed in germ cell tumors with constitutive overexpression altering human germline differentiation and proliferation. *Stem Cell Res.* *27*, 46–56.
- Gkoutela, S., Li, Z., Vincent, J.J., Zhang, K.X., Chen, A., Pellegrini, M., and Clark, A.T. (2013). 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.
- Gkoutela, S., Zhang, K.X., Shafiq, T.A., Liao, W.W., Hargan-Calvopina, J., Chen, P.Y., and Clark, A.T. (2015). DNA demethylation dynamics in the human prenatal germline. *Cell* *161*, 1425–1436.
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* *32*, 2847–2849.



- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The transcriptome and DNA methylation landscapes of human primordial germ cells. *Cell* *161*, 1437–1452.
- Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013). Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem Cell Reports* *1*, 518–531.
- Hargan-Calvopina, J., Taylor, S., Cook, H., Hu, Z., Lee, S.A., Yen, M.R., Chiang, Y.S., Chen, P.Y., and Clark, A.T. (2016). Stage-specific demethylation in primordial germ cells safeguards against precocious differentiation. *Dev. Cell* *39*, 75–86.
- He, J., Wang, Y., and Li, Y.L. (2007). Fibroblast-like cells derived from the gonadal ridges and dorsal mesenteries of human embryos as feeder cells for the culture of human embryonic germ cells. *J. Biomed. Sci.* *14*, 617–628.
- Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* *555*, 392–396.
- Hua, J., Yu, H., Liu, S., Dou, Z., Sun, Y., Jing, X., Yang, C., Lei, A., Wang, H., and Gao, Z. (2009). Derivation and characterization of human embryonic germ cells: serum-free culture and differentiation potential. *Reprod. Biomed. Online* *19*, 238–249.
- Irie, N., Weinberger, L., Tang, W.W., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. *Cell* *160*, 253–268.
- Kagiyada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.* *32*, 340–353.
- Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., Moritoki, Y., Kohri, K., Kimura, H., Yamamoto, T., et al. (2015). Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. *Cell Stem Cell* *16*, 517–532.
- Leitch, H.G., Nichols, J., Humphreys, P., Mulas, C., Martello, G., Lee, C., Jones, K., Surani, M.A., and Smith, A. (2013). Rebuilding pluripotency from primordial germ cells. *Stem Cell Reports* *1*, 66–78.
- Liu, S., Liu, H., Pan, Y., Tang, S., Xiong, J., Hui, N., Wang, S., Qi, Z., and Li, L. (2004). Human embryonic germ cells isolation from early stages of post-implantation embryos. *Cell Tissue Res.* *318*, 525–531.
- Matsui, Y., Zsebo, K., and Hogan, B.L. (1992). Derivation of pluripotent embryonic stem cells from murine primordial germ cells in culture. *Cell* *70*, 841–847.
- Mooijman, D., Dey, S.S., Boisset, J.-C., Crosetto, N., and van Oudenaarden, A. (2016). Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* *34*, 852–856.
- Ohta, H., Kurimoto, K., Okamoto, I., Nakamura, T., Yabuta, Y., Miyauchi, H., Yamamoto, T., Okuno, Y., Hagiwara, M., Shirane, K., et al. (2017). In vitro expansion of mouse primordial germ cell-like cells recapitulates an epigenetic blank slate. *EMBO J.* *36*, 1888–1907.
- Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., et al. (2014). Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet.* *10*, e1004868.
- Oliveros-Etter, M., Li, Z., Nee, K., Hosohama, L., Hargan-Calvopina, J., Lee, S.A., Joti, P., Yu, J., and Clark, A.T. (2015). PGC reversion to pluripotency involves erasure of DNA methylation from imprinting control centers followed by locus-specific re-methylation. *Stem Cell Reports* *5*, 337–349.
- Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., and Clark, A.T. (2016). Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* *18*, 323–329.
- Prokopuk, L., Stringer, J.M., Hogg, K., Elgass, K.D., and Western, P.S. (2017). PRC2 is required for extensive reorganization of H3K27me3 during epigenetic reprogramming in mouse fetal germ cells. *Epigenetics Chromatin* *10*, 7.
- Resnick, J.L., Bixler, L.S., Cheng, L., and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature* *359*, 550–551.
- Rooijers, K., Markodimitraki, C.M., Rang, F.J., de Vries, S.S., Chia-lastri, A., de Luca, K.L., Mooijman, D., Dey, S.S., and Kind, J. (2019). Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat. Biotechnol.* *37*, 766–772.
- Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., et al. (2015). Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* *17*, 178–194.
- Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Shiraki, N., Takakuwa, T., et al. (2016). The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Dev. Cell* *39*, 169–185.
- Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* *48*, 849–862.
- Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* *278*, 440–458.
- Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* *134*, 2627–2638.
- Sen, M., Mooijman, D., Boisset, J.-C., Chialastri, A., Popovic, M., Heindryckx, B., de Sousa Lopes, S.M.C., Dey, S.S., and van Oudenaarden, A. (2019). Strand-specific single-cell methylomics reveals distinct modes of DNA demethylation dynamics during early mammalian development. *bioRxiv* <https://doi.org/10.1101/804526>.
- Shablott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., and Gearhart, J.D.



- (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc. Natl. Acad. Sci. U S A* *95*, 13726–13731.
- Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature* *511*, 611–615.
- Sybirna, A., Tang, W.W.C., Dietmann, S., Gruhn, W.H., and Surani, A.M. (2019). A critical but divergent role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. *bioRxiv* <https://doi.org/10.1101/563072>.
- Tang, W.W., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., and Surani, M.A. (2015). A unique gene regulatory network resets the human germline epigenome for development. *Cell* *161*, 1453–1467.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* *25*, 1105–1111.
- Turnpenny, L., Brickwood, S., Spalluto, C.M., Piper, K., Cameron, I.T., Wilson, D.I., and Hanley, N.A. (2003). Derivation of human embryonic germ cells: an alternative source of pluripotent stem cells. *Stem Cells* *21*, 598–609.
- Turnpenny, L., Spalluto, C.M., Perrett, R.M., O'Shea, M., Hanley, K.P., Cameron, I.T., Wilson, D.I., and Hanley, N.A. (2006). Evaluating human embryonic germ cells: concord and conflict as pluripotent stem cells. *Stem Cells* *24*, 212–220.
- Vincent, J.J., Huang, Y., Chen, P.Y., Feng, S., Calvopina, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., et al. (2013). Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell* *12*, 470–478.
- von Meyenn, F., Berrens, R.V., Andrews, S., Santos, F., Collier, A.J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P.J., and Reik, W. (2016). Comparative principles of DNA methylation reprogramming during human and mouse in vitro primordial germ cell specification. *Dev. Cell* *39*, 104–115.
- Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* *10*, 232.
- Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* *16*, 284–287.
- Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M., Gumucio, D.L., et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. *Nature* *573*, 421–425.

Chapter 5

Divergent roles for KLF4 and TFCP2L1 in naive ground state pluripotency and human primordial germ cell development

Divergent roles for KLF4 and TFCEP2L1 in Naive Ground State Pluripotency and Human Primordial Germ Cell Development

Hancock, G.^{1,2,3}, Liu, W.⁴, Peretz, L.¹, Chen, D.^{1,3*}, Gell, JJ.^{3,5,6**}, Collier, AJ.⁷, Zamudio, JR.¹, Plath, K.^{2,3,7}, Clark, AT^{1,2,3,6}

1. Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, California, USA
2. Molecular Biology Institute, University of California, Los Angeles, California, USA
3. Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, California, USA
4. Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Hangzhou 310058, P. R. China
5. David Geffen School of Medicine, Department of Pediatrics, Division of Hematology-Oncology, Los Angeles, CA, 90095, USA
6. Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California, USA
7. Department of Biological Chemistry, David Geffen School of Medicine, Los Angeles, California, USA

*Present address for D Chen Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Hangzhou 310058, P. R. China

** Present address for JJ Gell Department of Pediatrics, University of Connecticut, Farmington, Connecticut, USA, Connecticut Children's, Hartford, Connecticut, USA

Corresponding author, Amander Clark, 615 Charles E Young Drive South, Los Angeles, CA 90095, e.mail: clarka@ucla.edu

Keywords: Primordial Germ Cells, PGCs, hPGCs, Pluripotency, KLF4, TFCEP2L1, stem cells

Summary

During embryo development, human primordial germ cells (hPGCs) express a naive gene expression program with similarities to pre-implantation naive epiblast (EPI) cells and naive human embryonic stem cells (hESCs). Previous studies have shown that TFAP2C is required for establishing naive gene expression in these cell types, however the role of additional naive transcription factors in hPGCs is not known. Here, we show that unlike TFAP2C, the naive transcription factors KLF4 and TFCP2L1 are not required for the induction of hPGC-like cells (hPGCLCs) from hESCs, and they have no role in establishing and maintaining the naive-like gene expression program with extended time in culture. Taken together, our results suggest a model whereby the molecular mechanisms that drive naive gene expression in hPGCs/hPGCLCs are distinct from those in the naive EPI/hESCs.

Introduction

The first three weeks of human embryo development are characterized by extensive developmental progression from a relatively simple blastocyst at the end of week one post-fertilization (pf), to a complex conceptus embedded in the wall of the uterus in which the nascent embryonic sac forms, the embryonic disc acquires asymmetry and gastrulation occurs. In addition, during this time, primordial germ cells (PGCs) are specified and the extraembryonic compartment increases in cell number and complexity in order to facilitate the patterning of the developing embryo and sustain the pregnancy. As the embryo implants into the uterus, the pluripotent cells transition through key pluripotent states. Namely, the inner cell mass (ICM) cells of the blastocyst progress into pre-implantation and early post implantation naive and formative EPI cells, followed by a transition into post-implantation primed EPI cells (Smith,

2017). As the primed EPI undergoes gastrulation, the pluripotent gene expression program is repressed as somatic cell fates are established. In contrast to developing somatic cells, newly specified hPGCs express a naive-like pluripotent gene expression program similar to the ICM/naive-EPI, including expression of diagnostic transcription actors KLF4, TFAP2C and TFCP2L1 (Tang et al., 2015; Chen et al., 2018). This has led to the hypothesis that mechanisms controlling naive pluripotent gene expression in the ICM/EPI are similar to those responsible for naive-like gene expression in hPGCs.

The hPGC lineage is specified in the early post-implantation embryo during the pluripotent transition from naive to primed (for a comprehensive review of PGC specification see Hancock et al., 2021). Using the mouse as a model, PGCs are specified from embryonic (E) day 6.25 epiblast cells *in vivo* (Ohinata et al., 2005), and formative epiblast-like cells (EpiLCs) *in vitro* (Hayashi et al., 2011). Once specified, mouse PGCs exhibit a pluripotent-like gene expression program that can be reset to self-renewing pluripotency *in vitro* through the derivation of embryonic germ cells (EGCs) from *in vitro* cultured PGCs (Leitch et al., 2013). In hPGCs, the naive-like pluripotent gene expression program is maintained until around week 10 pf (Li et al., 2017). After which the hPGCs heterogeneously repress the naive-like program while differentiating into oogonia and pro-spermatogonia. Although hPGCs express diagnostic naive transcription factors that are required for naive pluripotent self-renewal, hPGCs are considered unipotent as their only fate is to differentiate into gametes. Failure to repress the naive gene expression program in hPGCs is speculated to be associated with development of germ cell tumors (Gell et al., 2018; Schmoll, 2002), and transformation of hPGCs into EGCs *in vitro* (Shamblott et al., 1998). However unlike EGCs in the mouse which have a similar gene

expression program to ESCs (Sharova et al., 2007), hEGCs do not exhibit robust self-renewal *in vitro* unless transduced with lentiviruses expressing OCT4 and SOX2 which are speculated to supplement the effect of KLF4 and cMYC (Bazley et al., 2015). Taken together, understanding the acquisition and control of naive-like pluripotent gene expression in hPGCs may provide critical insights into the similarities and differences between functional pluripotency in the EPI, the unipotency of hPGCs in the embryo, and the transformation of hPGCs into germ cell tumors *in vivo* and EGCs *in vitro*.

In the last decade a number of groups have developed media formulations to stabilize the self-renewing state of naive, formative and primed human pluripotent cells (Chan et al., 2013; Gafni et al., 2013; Hanna et al., 2010; Kinoshita et al., 2021; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014). The naive state is modelled *in vitro* by culturing hESCs or human induced pluripotent stem cell (hiPSCs) in a media called t2i/L/Go (t2iLGo) or 5iLA/F (5iLAF) (Takashima et al., 2014; Theunissen et al., 2014). Naive cells are refractory to differentiation *in vitro*, and instead require Wnt inhibition to dismantle the naive gene expression program in order to establish the differentiation-competent formative state of pluripotency (Rostovskaya et al., 2019). Culture of human cells in the formative pluripotent state requires transitioning naive hESCs or ICM explants into low levels of Activin, the Tankyrase inhibitor XAV939 and a retinoic acid receptor inverse agonist (*AloXR*) (Kinoshita et al., 2021). Finally primed human pluripotency is stabilized under conventional culture conditions in KSR/FGF2 and these cells are poised to differentiate rapidly (Thomson et al., 1998).

Modelling hPGC specification *in vitro* involves differentiating hPGC-like cells (hPGCLCs) from hESCs/hiPSCs (Irie et al., 2015; Sasaki et al., 2015; Tang et al., 2015). In

humans, the transcription factor TFAP2C is required to both establish the naive-like gene expression in hPGC-like cells (Chen et al., 2018) and to open naive enhancers to promote expression of the naive transcriptome in hESCs (Pastor et al., 2018). Similar to TFAP2C, KLF family members also regulate transcription in naive hESCs, with KLF4-KLF17 transcription factors responsible for binding young transposable elements (TE) at sites called TE Enhancers (TEEnhancers) to drive expression of neighboring genes (Pontis et al., 2019). The transcription factor that functions synergistically with KLF4 to maintain naive pluripotency in t2iLGo is Transcription Factor CP2-like protein 1 (TFCP2L1) (Takashima et al., 2014). TFCP2L1 and KLF4 function downstream of Wnt/ β -catenin to induce and maintain the naive pluripotent state in the mouse (Qiu et al., 2015). However, it is unclear whether KLF4 and TFCP2L1 regulate the specification and establishment of naive-like gene expression hPGCLCs.

In order to address this, we characterized KLF4 and TFCP2L1 expression with induction of hPGCLCs from hESCs, and used CRISPR/Cas9 gene editing to address the function of these transcription factors in hPGCLC induction and hESC reversion to the naive state in 5iLAF.

Results

KLF4 protein is expressed before TFCP2L1 during hPGCLC induction from hESCs.

To characterize KLF4 and TFCP2L1 we differentiated hPGCLCs from hESCs through an incipient mesoderm-like cell (iMeLC) intermediate (Sasaki et al., 2015). Fluorescence activated cell sorting (FACS) was used to isolate hPGCLCs at day 4 (D4) of differentiation using conjugated antibodies that recognize Integrin alpha-6 (ITGA6) and epithelial cell adhesion molecule (EPCAM) which separates the hPGCLCs from the differentiating somatic cells (Chen et al., 2017; Sasaki et al., 2015). Semi quantitative reverse transcriptase polymerase chain

reaction (RT-PCR) was performed on the hESCs and hPGCLCs showing that hPGC markers *PRDM1*, *SOX17* and *TFAP2C* are significantly up regulated in hPGCLCs relative to undifferentiated hESCs as previously reported (Irie, et al., 2015, Sasaki et al., 2015; Kojima et al., 2017) and this is also the case for *KLF4* and *TFCP2L1* (Fig. 1a).

Next, we performed immunofluorescence of aggregates at D4 to evaluate KLF4 and TFCP2L1 protein expression in the newly induced hPGCLCs. We discovered that KLF4 is co-expressed with TFAP2C/PRDM1 in hPGCLCs, while TFCP2L1 protein was not detected at this stage (Fig. 1b). To verify this result, we evaluated TFCP2L1 and KLF4 protein expression in hPGCLCs differentiated from UCLA2 (46, XY) hESCs (Supplementary Fig. 1a) and show again that KLF4 protein but not TFCP2L1 is detectable at D4 in hPGCLCs.

Given that both KLF4 and TFCP2L1 protein were previously identified in week 7 genital ridge BLIMP1+ hPGCs (Tang et al., 2015), we expanded upon this finding to better understand KLF4 and TFCP2L1 expression in cKIT+ hPGCs *in vivo*. To achieve this, we examined either KLF4 or TFCP2L1 protein in cKIT+ hPGCs from day 74 (D74) to D140 pf in both male and female gonads (Fig. 1c). Although hPGC differentiation into oogonia and spermatogonia are initiated at around week 10 pf, cKIT+ hPGCs persist in the developing fetal gonad well into the second trimester (Gkountela et al., 2013). In 100% of human fetal testis (n=6) and fetal ovary (n=6) samples evaluated between D74-D140 we show that KLF4 and TFCP2L1 protein are expressed in the nucleus of cKIT+ hPGCs (Supplementary figure 1b). In addition, we discovered that KLF4 protein was expressed in rare cKIT negative cells, which were outside the seminiferous tubules of the fetal testes. KLF4+ somatic gonadal cells were also observed in a recently published 10x Genomics single cell RNA-Seq data set of the human fetal testis and ovary (Chitiashvili et al., 2020), confirming that KLF4 is expressed by both hPGCs and somatic

cells in the embryonic and fetal gonad. In contrast, TFCEP2L1 protein was specific to cKIT+ hPGCs in both males and females (Fig. 1c). Taken together, KLF4 is upregulated with hPGCLC induction from hESCs and both proteins are expressed throughout hPGC development in vivo with KLF4 also expressed in gonadal somatic cells.

KLF4 and TFCEP2L1 are required to reset naive pluripotency in 5iLAF *in vitro*

To investigate the role of KLF4 and TFCEP2L1 in hPGC development using the hPGCLC model, we created null mutant hESC sublines for each gene following picking and expanding CRISPR-Cas9 gene edited colonies. Specifically, we produced two KLF4 mutant hESC sublines in the UCLA1 (46, XX) hESC line (#35 and #43), and one in the H1 OCT4-GFP line (46, XY). For TFCEP2L1, we produced two mutant sublines (#3 and #7) in the UCLA1 hESC line. Mutations in each of the sublines were confirmed by genotyping and Sanger sequencing of the isolated DNA (Supplementary Fig. 2a, b).

Previous studies showed that knockdown of KLF4 and TFCEP2L1 in the naive media t2iLGo disrupts naive colony formation (Takashima et al., 2014). In order to evaluate the role of KLF4 and TFCEP2L1 in under 5iLAF naive culture conditions, we reverted control and mutant hESC sublines to the naive state using 5iLAF and used FACS to quantify the percentage of CD75+ cells (naive) as well as CD24+ (primed) cells (Fig. 2a) (Collier et al., 2017). Using this approach, we found a significant reduction in the CD75+ / CD24- naive population relative to controls following naive reversion (Fig. 2b and supplementary figure 2c). As additional confirmation, the null mutant phenotype was shown by immunofluorescence for KLF4 and TFCEP2L1 proteins, which were detectable in the dome shaped colonies of the control sublines but absent in the residual surviving mutant cells after naive reversion (Fig. 2c,d). To verify that

the colonies in the control cultures were naive, we performed immunofluorescence for the naive marker KLF17 as well as the pan-pluripotent transcription factor OCT4 (Fig. 3c,d). These results show that KLF17 and OCT4 are expressed in the dome shaped colonies of the control cultures, whereas in the mutant sublines, KLF17 and OCT4 proteins were sporadically expressed in random cells not organized into colonies. Together these findings show that CRISPR/Cas9 gene editing of KLF4 and TFCP2L1 loci result in null mutations which cause a significant reduction in the capacity of primed hESCs to revert to the naive state in the media 5iLAF.

KLF4 and TFCP2L1 are not required for naive pluripotent gene expression in hPGCLCs

Next, we evaluated the role of KLF4 in hPGC development by differentiating hPGCLCs from control and mutant hESC sublines. To quantify hPGCLC induction, we performed FACS for ITGA6/EPCAM at D4 (Fig. 3a-c) and show that all three KLF4 mutant sublines (UCLA1 #35, #43; H1 #39) produce ITGA6/EPCAM double positive hPGCLCs at percentages equivalent to their respective controls. Similarly, hPGCLC induction in the TFCP2L1 mutant sublines (#3, #7) was also unaffected relative to control (Fig. 3d and e). This suggests that hPGCLC induction from primed hESCs does not require KLF4 or TFCP2L1.

Even though hPGCLC specification had occurred in the absence of KLF4, it is conceivable that naive gene expression in the resulting hPGCLCs is de-regulated. To address this, we performed RNA-Seq of ITGA6/EPCAM FACS-isolated KLF4 null mutant (H1 #39) and H1 control hPGCLCs at D4 of differentiation and show that the KLF4 null mutant hPGCLC transcriptome is overall very similar to that of control cells (Supplementary figure 3b). The rare exceptions include *KLF4*, (Supplementary Fig. 3c), which we also confirmed was not expressed at the protein level in hPGCLCs (Supplementary Fig. 3a). The remaining differentially

expressed genes (DEGs) were unrelated to naive pluripotent gene expression (Supplementary figure 3d). To highlight the similarity in gene expression between KLF4 mutant and control hPGCLCs, we displayed diagnostic genes of hPGC development, naive and primed pluripotency, shared pluripotency and somatic cell differentiation on a heatmap (Supplementary Figure 3d). This display highlights the similarity in naive, primed and shared pluripotent gene expression across the data sets, as well as a robust early hPGC gene expression program similar to hPGCs *in vivo* (Supplementary Figure 3d). Taken together, these results indicate that while KLF4 is highly upregulated upon hPGCLC specification, it is not functionally required for induction of hPGCLCs from hESCs and is not required for establishing the transcriptional state of naive-like pluripotency in hPGCLCs.

TFCP2L1 and KLF4 are not required for hPGCLC maintenance or proliferation

Next, we used the extended culture system (Gell et al., 2020) to evaluate whether KLF4 and TFCP2L1 proteins regulate hPGCLC survival and proliferation after induction of the initial hPGCLC population in the aggregates. To achieve this, we used FACS to isolate D4 hPGCLCs with ITGA6/EPCAM, and cultured the hPGCLCs for an additional 21 days in extended culture (D4C21) (Fig. 4a). During the course of extended culture hPGCLCs do not revert to EGCs (Gell et al., 2020), and maintain germ cell identity as shown by co-expression of SOX17, PRDM1, and TFAP2C (Fig. 4b). Furthermore, by D10 of extended culture, TFCP2L1 protein is detectable in PRDM1/TFAP2C double positive hPGCLCs (Fig. 4c). This result indicates that TFCP2L1 becomes stably expressed after KLF4 in hPGCLC development.

Given that both TFCP2L1 and KLF4 protein are expressed in hPGCLCs during extended culture, we sought to evaluate the role of KLF4 and TFCP2L1 in hPGCLC proliferation by

exposing cells to Edu at D4C21, and calculating the percentage of Edu+ cells in the PRDM1/TFAP2C double positive hPGCLC colonies. Under control conditions, we show that approximately 30% of PRDM1/TFAP2C positive hPGCLCs incorporate Edu, and are therefore in S-phase of the cell cycle during the 4-hour window of Edu exposure (Fig. 5a-d). Analysis of Edu incorporation in hPGCLCs derived from the TFCP2L1 mutant sublines (#3 and #7), and the KLF4 mutant sublines (#35 and #43) shows that the percentage of Edu+ cells is equivalent to the controls. Similarly, the fraction of PRDM1/TFAP2C hPGCLC colonies normalized to plating density in control and mutants is unaffected (Supplementary figure 4 a-c). Together these results show that KLF4 and TFCP2L1 protein are expressed in hPGCs and hPGCLCs; however, each is dispensable for hPGCLC induction, survival, proliferation and establishment of naive-like pluripotency.

Discussion

Our results highlight notable differences between the mechanism responsible for establishing naive pluripotent gene expression in naive EPI/hESCs and hPGCs/hPGCLCs. In previous studies, it was shown that TFAP2C binds to and opens naive enhancers to regulate the establishment and maintenance of naive ground state pluripotency in 5iLAF and t2iLGo, as well as the specification and reacquisition of naive-like ground state pluripotency in hPGCLCs (Chen et al., 2018; Chen et al., 2019; Pastor et al., 2018). In the current study, we evaluated two additional regulators of naive hESC pluripotency, KLF4 and TFCP2L1, which similar to TFAP2C regulate self-renewal of naive cells in t2iLGo (Takashima et al., 2014). Here we show that KLF4 and TFCP2L1 also regulate naive pluripotency when reverting primed cells to the naive state in the media 5iLAF. However, unlike TFAP2C, our data showed that KLF4 and

TFCP2L1 are not required for hPGCLC induction or maintenance in culture. This suggests that even though a common naive gene expression program can be identified in naive hESCs and the hPGCs, only TFAP2C has a role in establishing the naive-like state of pluripotency in both hPGCLCs and hESCs, whereas KLF4 and TFCP2L1 function only in t2iLGo and 5iLAF naive hESCs.

Mechanistically, TFCP2L1 has been shown to target the KLF4 promoter, and to also bind KLF4 protein (Wang et al., 2019), highlighting that these transcription factors function in combinatorial roles in naive hESCs. In the current study we observed KLF4 protein at the time of hPGCLC specification whereas TFCP2L1 protein was expressed later, suggesting that TFCP2L1 protein most likely does not act upstream of KLF4 in hPGCLC induction. This uncoupling of TFCP2L1 protein expression from KLF4 at the time of hPGCLC specification may be necessary to avoid re-establishment of functional naive pluripotency, and therefore a protective mechanism to reduce the risk of germ cell tumors. Indeed the TFCP2L1 locus is in a haplotype block associated with testicular germ cell tumors (Wang et al., 2017).

Our results show that neither TFCP2L1 nor KLF4 regulate proliferation and survival of the hPGCLC population in extended culture. Success of the hPGCLC extended culture system depends on addition of cAMP agonists including Forskolin and Rolipram (Gell et al., 2020), which increase the intracellular content of cAMP. This is important because in placental cell lines, the expression of TFCP2L1 protein is associated with cAMP signaling (Henderson et al., 2008). Collectively this could suggest a relationship between TFCP2L1 protein expression and intracellular cAMP which could be investigated in future studies.

Our studies indicate that KLF4 had no role in early hPGC development, either at the time of hPGC specification or during extended culture. In the mouse, KLF4 is reported to function

redundantly with other KLF family members including KLF2 and KLF5 to regulate naive pluripotency by co-binding critical pluripotent transcription factors including POU5f1, SOX2, NANOG, ESRRB (Jiang et al., 2008). Our RNA-Seq of hPGCLCs, together with previously published data sets of FACS isolated hPGCs (Chen et al., 2019; Tang et al., 2015) indicates that *KLF5 KLF11, KLF13* and *KLF16* are all highly expressed in hPGCLCs/hPGCs, whereas *KLF2* expression is below the limit of detection. Therefore, it could be hypothesized that similar to the mouse, other KLF family members substitute for KLF4 in regulating hPGC development. However, none of the KLF family members (including KLF17) were upregulated as a consequence of KLF4 knockout in hPGCLCs.

In summary, the requirement for KLF4 and TFAP2L1 in establishing naive pluripotency in human pluripotent stem cells but not in hPGCLCs likely highlights a critical mechanistic difference in these closely related embryonic cell types, and illustrates that establishment and maintenance of naive-like pluripotency in hPGCLCs is critically dependent on TFAP2C. Improved late-stage modeling of hPGC development including expression of gonadal hPGC markers, together with a more detailed understanding of naive-like transcription factors in germ cell development and pluripotency, collectively will provide a better understanding of causes of infertility and germ cell tumors in the context of healthy embryonic development.

Experimental Procedures

Human fetal samples

All prenatal gonads 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. All consented material was donated anonymously and carried no personal identifiers. Therefore the use of the de-identified fetal tissue at UCLA was deemed exempt by the UCLA IRB under 45 CFR 46.102(f). Developmental age was documented by BDRL as days post fertilization using prenatal intakes, foot length, Streeter's Stages and crown-rump length. All prenatal gonads documented with birth defect or chromosomal abnormality were excluded from this study.

Human ESC culture

The hESC lines in this study are as follows: UCLA1 (46, XX), UCLA2 (46, XY)(Diaz Perez et al., 2012), UCLA8 (46, XX)(Chen et al., 2017), and H1 OCT4-GFP (H1) (46, XY)(Gkountela et al., 2015). All hESCs were cultured on mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) and split every 7 days using Collagenase type IV (GIBCO, 17104-019). hESC media was comprised of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100mM L-Glutamine (GIBCO,25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 55mM 2-Mercaptoethanol (GIBCO, 21985-023), 10ng/mL recombinant human FGF basic (Proteintech HZ1285), 1x Penicillin-Streptomycin (GIBCO, 15140-122), and 50ng/mL pri-mocin (InvivoGen, ant-pm-2) in DMEM/F12 media (GIBCO, 11330-032). All hESC lines used in this study are registered with the National Institute of Health Human Embryonic Stem Cell Registry and are available for research use with NIH funds. hESCs used in this study were routinely tested for mycoplasma (Lonza, LT07-418). All experiments were approved by the UCLA Embryonic Stem Cell Research Oversight Committee.

Realtime PCR

Undifferentiated hESCs, iMeLCs, D4 hPGCLCs (EPCAM/ITGA6), and D4 somatic cells (the EPCAM/ITGA6 negative cells) were re-suspended in 350 uL RLT buffer (QIAGEN) and RNA was extracted using RNeasy micro kit (QIAGEN). RNA was converted to cDNA using SuperScript® II Reverse Transcriptase (Invitrogen). Real-time quantitative PCR was performed using TaqMan® Universal PCR Master Mix (Applied Biosystems), and with Taqman probes detecting expression of GAPDH, TFAP2C, SOX17, PRDM1, KLF4, and TFCP2L1. Expression levels for genes of interest were normalized to housekeeping gene GAPDH in each cell type. To quantify relative expression in each cell type, expression levels was normalized to the expression level of hESCs. hPGCLCs were compared to hESCs in biological triplicate, and to iMeLCs and somatic cells in duplicate. P-values were calculated using two-tailed student T test.

Immunofluorescence

Aggregates collected at D4, and human fetal tissue samples were fixed in 4% PFA for 1 hour, washed twice for 15 minutes in PBS, stained with hemotoxylin, and mounted in histogel (Thermo Scientific). Samples were embedded into paraffin blocks and cut onto slides in 5 um-thick sections. Slides were deparaffinized and rehydrated through a series of xlene and ethanol series. For antigen retrieval, slides were heated to 95C in Tris-EDTA solution (10 mM Tris Base, 1 mM EDTA solution, .05% Tween-20, pH9.0). Sections were permeabilized (.05% Triton-100 in PBS) for 20 minutes and blocked in PBS containing 10% normal donkey serum for 1 hour. Primary antibodies incubated overnight at 4C. Antibodies included anti-TFAP2C (sc12762; 1:100), anti-SOX17 (GT15094; 1:100), anti-Blimp1 (9115S; 1:100), anti-KLF4 (AF3640; 1:100), anti-TFCP2L1 (AF5726; 1:100), and (cKIT A405 1:100). The next day, slides were

washed, blocked for an additional 30 minutes, and stained with secondary antibody for 1 hour in their corresponding species-specific secondary antibody. Secondary antibodies included donkey anti-mouse 488 IgG (715-546-150; 1:200), donkey anti-mouse 594 IgG (A21447; 1:200), donkey anti-rabbit 488 IgG (711-545-152; 1:200), donkey anti-rabbit 594 IgG (711-585-152; 1:200), donkey anti-rabbit 647 IgG (711-605-152; 1:200), donkey anti-goat 488 IgG (705-546-147; 1:200), donkey anti-goat 594 IgG KLF4, TFCP2L1 (705-586-147; 1:200), donkey anti-goat 647 IgG KLF4, TFCP2L1 (A21447; 1:200). Dapi (xxx; 1:1000) was added during secondary antibody incubation and samples were mounted in ProLong Gold antifade reagent (Invitrogen).

hPGCLCs at D4C10 or D4C21 on chamber slides, and 5iLAF cells split onto coverslips in culture at P3 were washed and fixed in 4% PFA for 10 minutes. Cells were washed, permeabilized, blocked, and stained as described above. Primary antibodies were anti-KLF17 (042649 1:200), and anti-OCT4 (sc-5279 1:100). Secondary antibodies were incubated for 30 minutes.

For Edu analysis, cells in culture were incubated with Edu for 4 hours fixed, and detected using Click-iT™ Edu Cell Proliferation Kit for Imaging, Alexa Flour™ 488 dye before permeabilization.

Image quantification

hPGCLCs in aggregates were quantified in IMARIS 8.1 (Bitplane). For KLF4 and quantification, we counted the percentage of cells that were KLF4+ in the TFCP2C, PRDM1 double-positive hPGCLC population. This was repeated in 3 cell lines. In human fetal samples,

we counted the percentage of cKIT positive hPGCs positive for KLF4 or TFCP2L1. This was repeated in 6 biological samples each of prenatal testes and ovaries.

In extended culture, total hPGCLC colonies in each well were counted and normalized to the number of cells plated for each sample. Each was performed with 3 technical replicates. For Edu quantification, the number of Edu positive cells in each hPGCLC colony was counted and Edu percentage of each hPGCLC colony was compared between the mutant and control-derived hPGCLCs. Error bars on graphs indicated standard error.

hESC mutants made by CRISPR/Cas9

To make null-mutations for KLF4 and TFCP2L1, pairs of gRNAs were designed to target the functionally important, most N-terminus coding region of each gene. Guides were designed using <https://zlab.bio/guide-design-resources>, and cloned into PX459 vector (Ran et al., 2013). Pairs of guides were designed approximately at a 3 kB distance from each other in the genome. 4-6 days before nucleofection, UCLA1 or H1-OCT4-GFP hESCs were purified from MEFs and transferred to matrigel (BD) in mTeSR media (stemcell tech). At 70% confluence, cells were electroporated with 4 ug of each gRNA pair was using P3 Primary Cell 4D-Nucleofector® X Kit according to the manufacturer's instructions (Lonza, V4XP-3024). 1 day following recovery, cells were dissociated with Accutase and replated on DR4 MEFs in a 6-well plate. Cells were treated with puromycin at a concentration of .35 ug/mL for 1 day. Once colonies emerge after 6-8 days, they are dissociated and plated at 10k and 50k densities in 10 cm plates. 48 colonies were picked when at the desired density after 10 days. Colonies were split in half after 4 days. To determine homozygous mutants, we genotyped genomic DNA and chose colonies with the

expected shorter band. Genotyping primers and gRNA sequences are listed in Supplementary table 1. To confirm bi-allelic mutations, mutant bands were cloned into Blunt-PCR-Cloning vector using Zero Blunt PCR Cloning Kit (ThermoFisher, K270020). 5-10 colonies from each band were picked and sequenced.

hPGCLC induction

hPGCLCs were induced as described previously (Chen et al., 2017). 1 hour before plating, 12-well plates are treated with human plasma fibronectin (Invitrogen). hESCs were washed and dissociated in 0.05% trypsin for 5 minutes, and quenched with MEF media. The MEFs were removed by plating in 10-cm cell culture dishes, twice for 5 minutes each. Purified hESCs were spun, filtered with 100um filter, and seeded at a density of 200k per 12-well in iMeLC media including 15% KSR (GIBCO, 10828-028), 1x NEAA (GIBCO, 11140-050), 0.1mM 2-Mercaptoethanol (GIBCO, 21985-023), 1x Penicillin-Streptomycin-Glutamine (GIBCO, 10378-016), 1mM sodium pyruvate (GIBCO, 11360-070), 50ng/mL Activin A (Peprotech, AF-120-14E), 3mM CHIR99021 (Stemgent, 04-0004), 10mM of ROCKi (Y27632, Stemgent, 04-0012-10), and 50ng/mL primocin in Glasgow's MEM (GMEM) (GIBCO, 11710-035). After 24 hours, cells were dissociated with trypsin, inactivated with trypsin inhibitor (Sigma), resuspended in PGLCC media, and plated in ultra-low cell attachment U-bottom 96-well plates (Corning) at a density of 3k cells/well. hPGCLC media is comprised of 15% KSR (GIBCO, 10828-028), 1x NEAA (GIBCO, 11140-050), 0.1mM 2-Mercaptoethanol (GIBCO, 21985-023), 1x Penicillin-Streptomycin-Glutamine (GIBCO, 10378-016), 1mM sodium pyruvate (GIBCO, 11360-070), 10ng/mL human LIF (Millipore, LIF1005), 200ng/mL human BMP4 (R&D systems, 314-BP),

50ng/mL human EGF (R&D systems, 236-EG), 10mM of ROCKi (Y27632, Stemgent, 04-0012-10), and 50ng/mL primocin in Glasgow's MEM (GMEM) (GIBCO, 11710-035).

Flow cytometry and fluorescence activated cell sorting

D4 aggregates were collected and dissociated with .05% trypsin for 10 minutes. The dissociated cells were stained with conjugated cell-surface antibodies for at least 15 minutes. Antibodies included anti-ITGA6-BV421 (BioLegend, 313624 1:60), anti-EPCAM-488 (BioLegend 324210; 1:60) for UCLA-derived lines and anti-ITGA6-488 (BioLegend, 313608 1:60), anti-EPCAM-APC (BioLegend 324208; 1:60) for H1-derived lines. After at least 15 minutes, cells were washed with FACS buffer (1% BSA in PBS), resuspended in FACS buffer with 7AAD (BD PharMingen 559925; 1:40). Single-cell suspensions of hESCs were used as single-color compensation controls and evaluated each with 7AAD, anti-ITGA6-BC421, and anti-EPCAM-488 for UCLA1-derived lines and anti-ITGA6-488, anti-EPCAM-APC, and unstained GFP-OCT4-expressing cells for the H1-OCT4-GFP derived lines. For gating controls, fluorescence-minus-one controls were made against each fluorophore, staining the residual cells from the aggregate supernatant with each antibody, minus its respective control. Gating was then established based on the absence of signal in the cell population of interest. Double-positive ITGA6, EPCAM cells were analyzed using an ARIA-H Fluorescence Activated Cell Sorter and sorted into either 350 uL of FR10 media or RLT buffer. Analysis was performed using FlowJo version 10.

For flow cytometry, as described previously (Collier et al., 2017), cells in 5iLAF at passage 3 were dissociated using accutase, passed through a 40um strainer (BD) and resuspended in FACS buffer at equal cell numbers. Conjugated antibodies including anti-CD75-APC (ThermoFisher

50-0759-41; 1:20), anti-CD24-BV421 (BD 562789 1:40), anti-CD90.2-APC-Cy7 (BioLegend 105327 1:20), and fixable live-dead-APC-Cy7 (Fisher 50-169-66) were diluted in staining buffer (BD 563794) and used to resuspend cell pellet, staining in the dark for at least 15 minutes. Live, non-mouse cells, through exclusion of CD90.2 positive mouse cells, were gated and analyzed for their percentage of CD75 positive, CD24 negative populations. Analyses was performed on an LSR Fortessa cytometer in duplicate.

Primed to Naive Reversion

Cells were reverted to the naive ground state in 5iLAF as described previously (Theunissen et al., 2014). At day 7, primed hESCs were dissociated into single cells with accutase and re-plated on MEFs in hESC media with Y27632 (Stemgent, 04-0012-10) at a density of 200k cells/well per 6-well plate. After one day, media was changed to 5iLAF media including a 50/50 mixture of DMEM/F12 (Gibco 11320-033) and Neurobasal (Gibco 21103-049), 1X N2 (Gibco 17502-048), 1X B27 (17504-044), 20 ng/mL rhLIF (Millipore LIF1005), 1 mM GlutaMAX (Gibco 35050-061), 1% NEAA (Gibco 11140-050), .1 mM 2-Mercaptoethanol (GIBCO, 21985-023), 1x Penicillin-Streptomycin (Gibco 15140-122), 50 ug/mL BSA (Gibco A10008-01), 1 mM PD0325901 (Stemgent 04-006-02), 1 mM IM-12 (BML-WN102-0005), .05 mM SB590885 (R&D 2650/10), 1 uM WH-4-023 (A Chemtek H620061), 10 uM Y-27632, 20 ng/mL Activin A (Peprotech AF-120-14E), 8 ng/mL FGF2 (Proteintech HZ1285), .50% KSR (Gibco 10828-028), and 1X primocin (Invitrogen (ant-pm-2)). Media is changed daily and cells are passaged every 5 days at a ratio between 1:1 and 1:3 until robust colonies emerge.

RNA sequencing library preparation and data analysis

Total RNA was extracted from H1-OCT4-GFP sorted D4 hPGCLCs using RNeasy micro kit (Qiagen 74004). Total RNA was reverse transcribed and cDNA was amplified using Nugen RNA-Seq System V2 (Nugen, 7102-32). DNA was extracted using MinElute PCR purification kit (Qiagen) and quantified using the Qubit dsDNA High-Sensitivity Kit (Life Technologies). Amplified cDNA was fragmented using Covaris S220 Focused-ultrasonicator. RNA-sequencing libraries were generated using Nugen Rapid Library Systems (Nugen 0320-32). Libraries were subjected to single-end 125 bp sequencing on HiSeq4000 with 6 indexed libraries per lane.

RNA sequencing analysis

Raw reads in qseq format obtained from the sequencer were first converted to fastq files with a customized perl script. Read quality was evaluated with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). High-quality reads were aligned to the hg19 human reference genome using Tophat (v 2.0.13) by using “-no-coverage-search” option, allowing up to two mismatches, and only keeping unique reads. The number of unique mappable reads was quantified by HTseq (v 0.5.4) under default parameters (Supplementary Table 2). Expression levels were determined by RPKM (reads per kilobase of exons per million aligned reads) in R using customized scripts. For RNAseq of published datasets, GSE76970 (Chen et al., 2017) and GSE93126 (Pastor et al., 2016), processed data of the raw read counts of each gene was utilized, with the same downstream analysis.

Extended culture

D4 hPGCLCs were cultured (C) for an additional 10 or 21 days as described previously (Gell et al., 2020). Sorted hPGCLCs were plated at densities ranging 200-3000 cells in either a chamber

well (D4C10) or 24-well (D4C21) in FR10 media. FR10 medium(Ohta et al., 2017) contains 10% KSR, 2.5% FBS (Thermo Fisher Scientific, SH3007003), 1× NEAA (Gibco, 11140-050), 1 mM sodium pyruvate (Gibco, 11360070), 2 mM L-glutamine (Gibco, 25030081), 0.1 mM 2-mercaptoethanol (Gibco, 21985-023), 1× penicillin streptomycin (Gibco, 15140-122), 100 ng/mL SCF (PeproTech, 250-03), 10 μM forskolin (Sigma, F6886), 10 μM rolipram (Sigma, R6520), and 50 ng/mL primocin in Glasgow's MEM (Gibco, 11710-035). For D4C21, hPGCLCs were dissociated at D4C10 using .05% trypsin for 3 minutes. Cells are spun down at 1.6 rpm for 5 minutes, carefully resuspended, and plated at a 1:2 ratio in chamber well. Media was replaced daily until readout.

Acknowledgements

The authors would like to thank Felicia Codrea, Jessica Scholes, and Jeffery Calimlim for FACS and Jinghua Tang for banking and culturing of the UCLA hESC lines. This work is supported by funds from an anonymous donor as well as NIH/NICHD R01 HD079546 (ATC), NIH/NIGMS P01 GM099134 (KP), and a Faculty Scholar grant from the Howard Hughes Medical Institute (KP). GH acknowledges the support of the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Training Program for supporting this work. Human fetal tissue research is supported by a grant to Ian Glass at the University of Washington Birth Defects laboratory 5R24HD000836-53. Human conceptus tissue requests can be made to bdrl@u.washington.edu.

Author Contributions

GH: Conceived and designed the experiments, performed the experiments, wrote the manuscript. LP, DC, and JG: performed the experiments. WL and JZ: performed the RNA sequencing analysis. AJC: designed the reversion and flow cytometry experiments. KP: designed the experiments and reviewed the manuscript. ATC: Conceived and designed the experiments, maintained all institutional compliances, received the funding, wrote the manuscript

Declaration of Interests

The authors declare no competing interests

References

Bazley, F.A., Liu, C.F., Yuan, X., Hao, H., All, A.H., De Los Angeles, A., Zambidis, E.T., Gearhart, J.D., Kerr, C.L., 2015. Direct Reprogramming of Human Primordial Germ Cells into

Induced Pluripotent Stem Cells: Efficient Generation of Genetically Engineered Germ Cells. *Stem Cells Dev* 24, 2634–2648. <https://doi.org/10.1089/scd.2015.0100>

Chan, Y.-S., Göke, J., Ng, J.-H., Lu, X., Gonzales, K.A.U., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., Ng, H.-H., 2013. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* 13, 663–675. <https://doi.org/10.1016/j.stem.2013.11.015>

Chen, D., Liu, W., Lukianchikov, A., Hancock, G.V., Zimmerman, J., Lowe, M.G., Kim, R., Galic, Z., Irie, N., Surani, M.A., Jacobsen, S.E., Clark, A.T., 2017. Germline competency of human embryonic stem cells depends on eomesodermin. *Biol. Reprod.* 97, 850–861. <https://doi.org/10.1093/biolre/iox138>

Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E., Clark, A.T., 2018. The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. *Cell Reports* 25, 3591-3602.e5. <https://doi.org/10.1016/j.celrep.2018.12.011>

Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W., Kellis, M., Clark, A., 2019. Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. *Cell Reports* 29, 4568-4582.e5. <https://doi.org/10.1016/j.celrep.2019.11.083>

Chitiashvili, T., Dror, I., Kim, R., Hsu, F.-M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K., Clark, A., 2020. Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nat Cell Biol* 22, 1436–1446. <https://doi.org/10.1038/s41556-020-00607-4>

Collier, A.J., Panula, S.P., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., Rugg-Gunn, P.J., 2017. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. *Cell Stem Cell* 20, 874-890.e7. <https://doi.org/10.1016/j.stem.2017.02.014>

Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., Clark, A.T., 2012. Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. *Hum Mol Genet* 21, 751–764. <https://doi.org/10.1093/hmg/ddr506>

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., Rais, Y., Shipony, Z., Mukamel, Z., Krupalnik, V., Zerbib, M., Geula, S., Caspi, I., Schneir, D., Shwartz, T., Gilad, S., Amann-Zalcenstein, D., Benjamin, S., Amit, I., Tanay, A., Massarwa, R., Novershtern, N., Hanna, J.H., 2013. Derivation of novel human ground state naive pluripotent stem cells. *Nature* 504, 282–286. <https://doi.org/10.1038/nature12745>

- Gell, J.J., Liu, W., Sosa, E., Chialastri, A., Hancock, G., Tao, Y., Wamaitha, S.E., Bower, G., Dey, S.S., Clark, A.T., 2020. An Extended Culture System that Supports Human Primordial Germ Cell-like Cell Survival and Initiation of DNA Methylation Erasure. *Stem Cell Reports*. <https://doi.org/10.1016/j.stemcr.2020.01.009>
- Gell, J.J., Zhao, J., Chen, D., Hunt, T.J., Clark, A.T., 2018. PRDM14 is expressed in germ cell tumors with constitutive overexpression altering human germline differentiation and proliferation. *Stem Cell Res* 27, 46–56. <https://doi.org/10.1016/j.scr.2017.12.016>
- Gkountela, S., Li, Z., Vincent, J.J., Zhang, K.X., Chen, A., Pellegrini, M., Clark, A.T., 2013. 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. <https://doi.org/10.1038/ncb2638>
- Gkountela, S., Zhang, K.X., Shafiq, T.A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y., Clark, A.T., 2015. DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell* 161, 1425–1436. <https://doi.org/10.1016/j.cell.2015.05.012>
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., Wang, W., Li, R., Yan, J., Zhi, X., Zhang, Y., Jin, H., Zhang, W., Hou, Y., Zhu, P., Li, J., Zhang, L., Liu, S., Ren, Y., Zhu, X., Wen, L., Gao, Y.Q., Tang, F., Qiao, J., 2015. The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* 161, 1437–1452. <https://doi.org/10.1016/j.cell.2015.05.015>
- Hancock, G.V., Wamaitha, S.E., Peretz, L., Clark, A.T., 2021. Mammalian primordial germ cell specification. *Development* 148. <https://doi.org/10.1242/dev.189217>
- Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., Jaenisch, R., 2010. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9222–9227. <https://doi.org/10.1073/pnas.1004584107>
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., Saitou, M., 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* 146, 519–532. <https://doi.org/10.1016/j.cell.2011.06.052>
- Henderson, Y.C., Frederick, M.J., Wang, M.T., Hollier, L.M., Clayman, G.L., 2008. LBP-1b, LBP-9, and LBP-32/MGR Detected in Syncytiotrophoblasts from First-Trimester Human Placental Tissue and Their Transcriptional Regulation. *DNA and Cell Biology* 27, 71–79. <https://doi.org/10.1089/dna.2007.0640>
- Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., Surani, M.A., 2015. SOX17 is a critical specifier of human primordial germ cell fate. *Cell* 160, 253–268. <https://doi.org/10.1016/j.cell.2014.12.013>

- Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S., Ng, H.-H., 2008. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* 10, 353–360. <https://doi.org/10.1038/ncb1698>
- Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J., Smith, A., 2021. Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell* 28, 453-471.e8. <https://doi.org/10.1016/j.stem.2020.11.005>
- Leitch, H.G., Nichols, J., Humphreys, P., Mulas, C., Martello, G., Lee, C., Jones, K., Surani, M.A., Smith, A., 2013. Rebuilding pluripotency from primordial germ cells. *Stem Cell Reports* 1, 66–78. <https://doi.org/10.1016/j.stemcr.2013.03.004>
- Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X., Zhu, X., Li, R., Yan, J., Wei, Y., Zhao, Y., Wang, W., Ren, Y., Yuan, P., Yan, Z., Hu, B., Guo, F., Wen, L., Tang, F., Qiao, J., 2017. Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* 20, 858-873.e4. <https://doi.org/10.1016/j.stem.2017.03.007>
- Ohinata, Y., Payer, B., O’Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., Saitou, M., Surani, M.A., 2005. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 436, 207–213. <https://doi.org/10.1038/nature03813>
- Ohta, H., Kurimoto, K., Okamoto, I., Nakamura, T., Yabuta, Y., Miyauchi, H., Yamamoto, T., Okuno, Y., Hagiwara, M., Shirane, K., Sasaki, H., Saitou, M., 2017. In vitro expansion of mouse primordial germ cell-like cells recapitulates an epigenetic blank slate. *EMBO J* 36, 1888–1907. <https://doi.org/10.15252/embj.201695862>
- Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., Clark, A.T., 2016. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell* 18, 323–329. <https://doi.org/10.1016/j.stem.2016.01.019>
- Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., Clark, A.T., 2018. TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat Cell Biol* 20, 553–564. <https://doi.org/10.1038/s41556-018-0089-0>
- Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T.W., Jaenisch, R., Trono, D., 2019. Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* 24, 724-735.e5. <https://doi.org/10.1016/j.stem.2019.03.012>

- Qiu, D., Ye, S., Ruiz, B., Zhou, X., Liu, D., Zhang, Q., Ying, Q.-L., 2015. Klf2 and Tfcp2l1, Two Wnt/ β -Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency. *Stem Cell Reports* 5, 314–322. <https://doi.org/10.1016/j.stemcr.2015.07.014>
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8, 2281–2308. <https://doi.org/10.1038/nprot.2013.143>
- Rostovskaya, M., Stirparo, G.G., Smith, A., 2019. Capacitation of human naïve pluripotent stem cells for multi-lineage differentiation. *Development* 146. <https://doi.org/10.1242/dev.172916>
- Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., Plath, K., 2017. Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* 20, 87–101. <https://doi.org/10.1016/j.stem.2016.10.006>
- Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., Nakamura, S., Sekiguchi, K., Sakuma, T., Yamamoto, Takashi, Mori, T., Woltjen, K., Nakagawa, M., Yamamoto, Takuya, Takahashi, K., Yamanaka, S., Saitou, M., 2015. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* 17, 178–194. <https://doi.org/10.1016/j.stem.2015.06.014>
- Schmoll, H.-J., 2002. Extragonadal germ cell tumors. *Annals of Oncology*, Educational Book of the 27th ESMO Congress, 18-22 October, 2002: Nice, France 13, 265–272. <https://doi.org/10.1093/annonc/mdf669>
- Sharova, L.V., Sharov, A.A., Piao, Y., Shaik, N., Sullivan, T., Stewart, C.L., Hogan, B.L.M., Ko, M.S.H., 2007. Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains. *Dev Biol* 307, 446–459. <https://doi.org/10.1016/j.ydbio.2007.05.004>
- Smith, A., 2017. Formative pluripotency: the executive phase in a developmental continuum. *Development* 144, 365–373. <https://doi.org/10.1242/dev.142679>
- Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., Reik, W., Bertone, P., Smith, A., 2014. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* 158, 1254–1269. <https://doi.org/10.1016/j.cell.2014.08.029>
- Tang, W.W.C., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., Surani, M.A., 2015. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* 161, 1453–1467. <https://doi.org/10.1016/j.cell.2015.04.053>

- Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., Lungjangwa, T., Imsoonthornruksa, S., Stelzer, Y., Rangarajan, S., D'Alessio, A., Zhang, J., Gao, Q., Dawlaty, M.M., Young, R.A., Gray, N.S., Jaenisch, R., 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* 15, 471–487. <https://doi.org/10.1016/j.stem.2014.07.002>
- Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., Jones, J.M., 1998. Embryonic stem cell lines derived from human blastocysts. *Science* 282, 1145–1147. <https://doi.org/10.1126/science.282.5391.1145>
- Vértesy, Á., Arindrarto, W., Roost, M.S., Reinius, B., Torrens-Juaneda, V., Bialecka, M., Moustakas, I., Ariyurek, Y., Kuijk, E., Mei, H., Sandberg, R., van Oudenaarden, A., Chuva de Sousa Lopes, S.M., 2018. Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells. *Nature Communications* 9, 1873. <https://doi.org/10.1038/s41467-018-04215-7>
- Wang, Xiaohu, Wang, Xiaoxiao, Zhang, S., Sun, H., Li, S., Ding, H., You, Y., Zhang, X., Ye, S.-D., 2019. The transcription factor TFCEP2L1 induces expression of distinct target genes and promotes self-renewal of mouse and human embryonic stem cells. *J. Biol. Chem.* 294, 6007–6016. <https://doi.org/10.1074/jbc.RA118.006341>
- Wang, Z., McGlynn, K.A., Rajpert-De Meyts, E., Bishop, D.T., Chung, C.C., Dalgaard, M.D., Greene, M.H., Gupta, R., Grotmol, T., Haugen, T.B., Karlsson, R., Litchfield, K., Mitra, N., Nielsen, K., Pyle, L.C., Schwartz, S.M., Thorsson, V., Vardhanabhuti, S., Wiklund, F., Turnbull, C., Chanock, S.J., Kanetsky, P.A., Nathanson, K.L., Testicular Cancer Consortium, 2017. Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. *Nat. Genet.* 49, 1141–1147. <https://doi.org/10.1038/ng.3879>
- Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., Tesar, P.J., Okada, J., Margaretha, L., Sperber, H., Choi, M., Blau, C.A., Treuting, P.M., Hawkins, R.D., Cirulli, V., Ruohola-Baker, H., 2014. Derivation of naive human embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4484–4489. <https://doi.org/10.1073/pnas.1319738111>

Figures and Legends

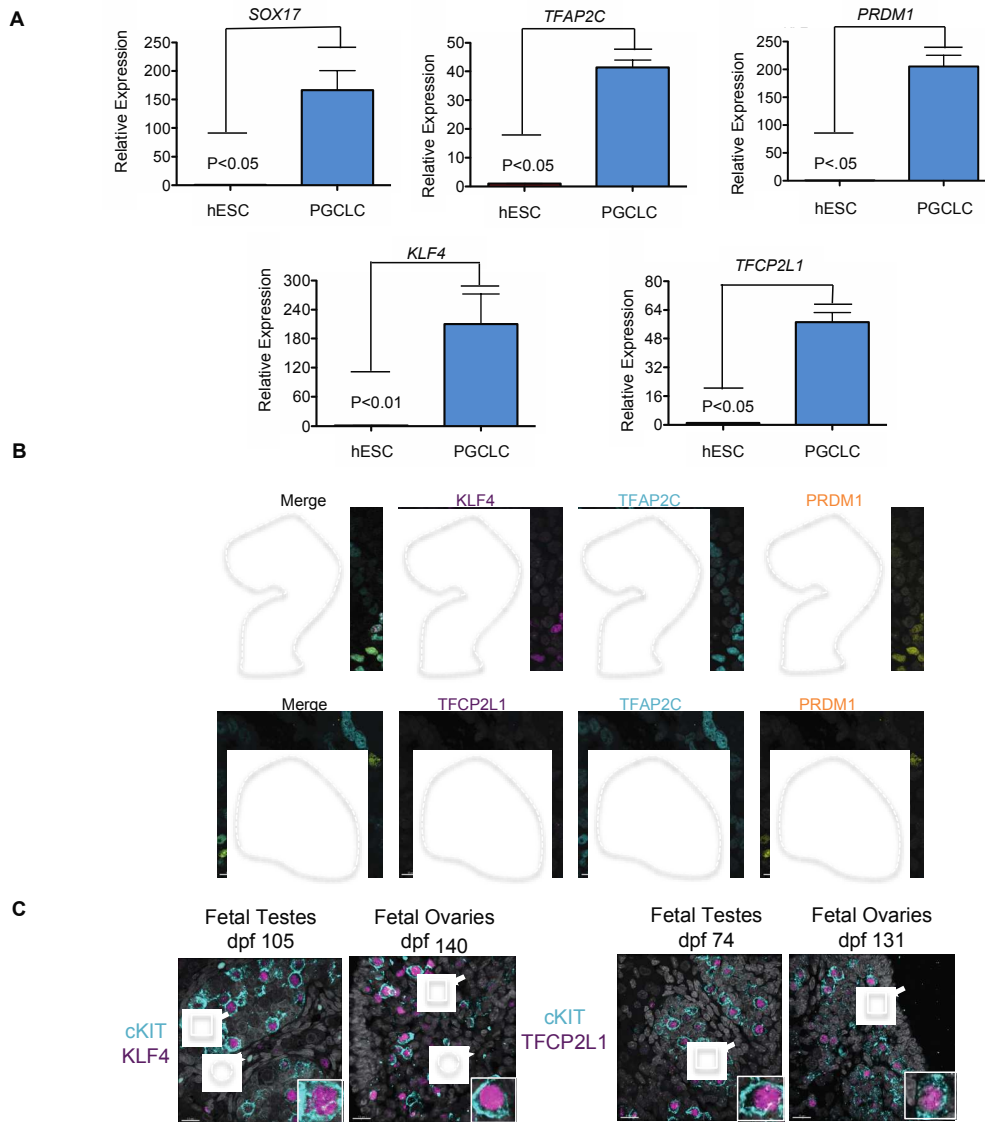


Figure 1. *KLF4* and *TFCP2L1* are dynamically expressed during hPGCLC specification. a. Quantitative real-time rt-PCR of *KLF4* and *TFCP2L1* in hESCs, and hPGCLCs from the UCLA1 hESC line at day 4 (D4) of differentiation. $n=3$ independent replicates of hESC and hPGCLCs. T-test was used to determine significance between these two groups. b. Representative immunofluorescence for *KLF4* and *TFCP2L1* in *TFAP2C*/*PRDM1* double positive hPGCLCs at D4 ($n=8$ aggregates of UCLA8). Scale bars show 15 microns. c. Representative immunofluorescence images of embryonic and fetal ovaries ($n=6$) and testes ($n=6$) at indicated days post fertilization (dpf): Arrows highlight double positive hPGCs, Arrowhead indicates single positive somatic cells.

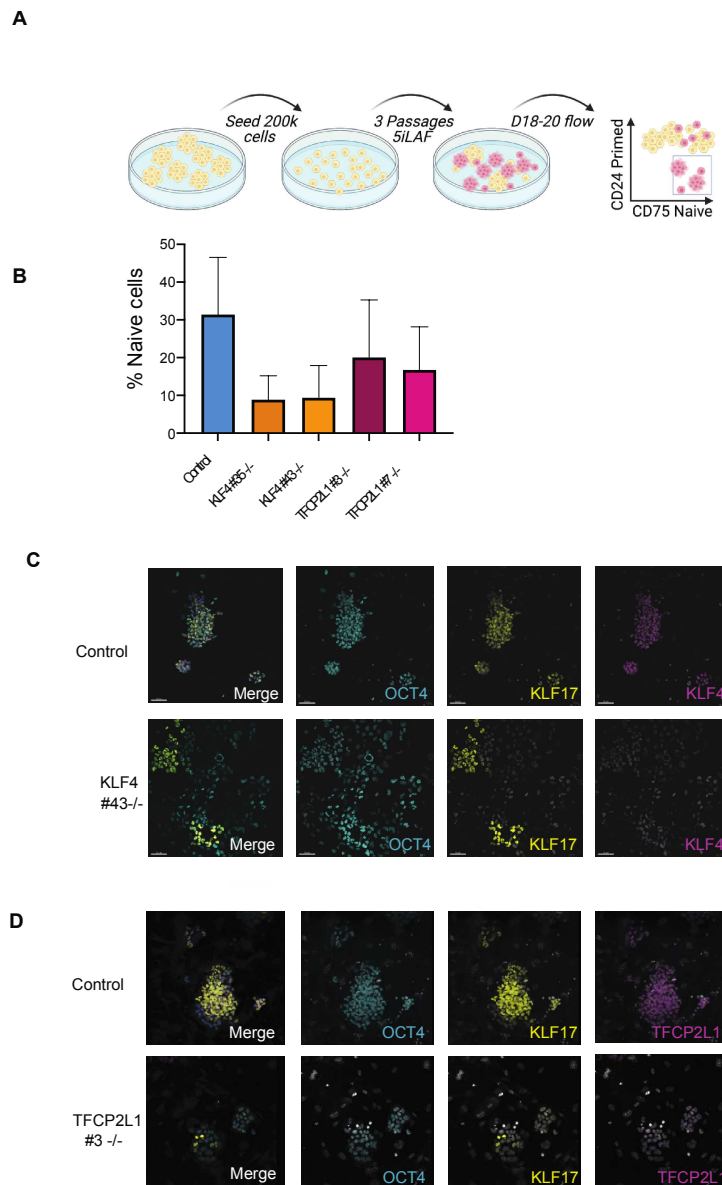


Figure 2. KLF4 and TFCP2L1 are required for naive pluripotency in 5iLAF.

a. Experimental schematic for reversion of hESCs from the primed to naive state. b. Quantification of CD75 positive, CD24 negative (naive) populations in the KLF4 and TFCP2L1 mutants and controls as evaluated by flow cytometry between D17-20 (n=3 each mutant cell line) c. Representative immunofluorescence image at passage 3 (D17-D19) reversions with KLF4 control and KO subline d. Representative immunofluorescence image at passage 3 (D17-D19) reversions with TFCP2L1 control and KO subline.

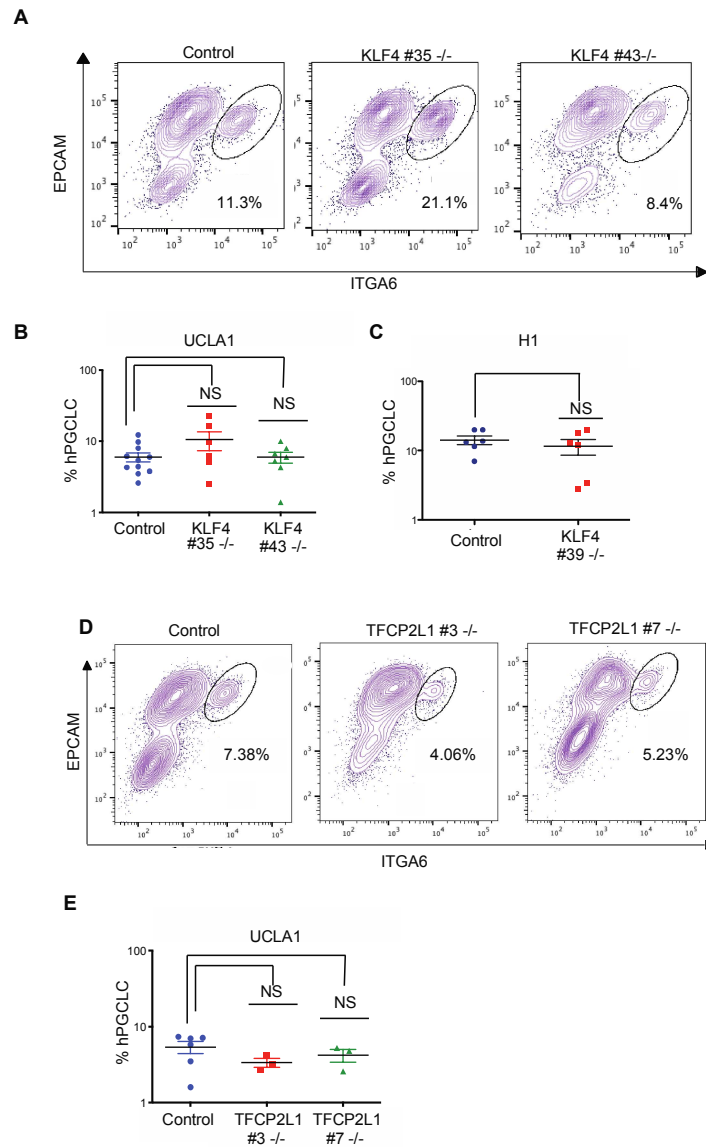


Figure 3. KLF4 and TFCP2L1 are not involved in hPGCLC induction from hESCs. a. Representative FACS plots showing EPCAM/ITGA6 double-positive hPGCLCs in control and KLF4 mutant aggregates at D4. b. Quantification of the percentage ITGA6/EPCAM hPGCLCs in control (n= 12 biological replicates) and KLF4 UCLA1 mutant sublines (n=6 and n=7 biological replicates respectively). c. Quantification of the percentage ITGA6/EPCAM hPGCLCs in control (n= 6 biological replicates) and KLF4 H1 mutant subline (n=6 biological replicates). d. Representative FACS plots showing EPCAM/ITGA6 double-positive hPGCLCs in TFCP2L1 control and mutant aggregates at D4. e. Quantification of the percentage ITGA6/EPCAM hPGCLCs in control (n= 8 biological replicates) and TFCP2L1 UCLA1 mutant sublines (n=3 and n=3 biological replicates respectively).

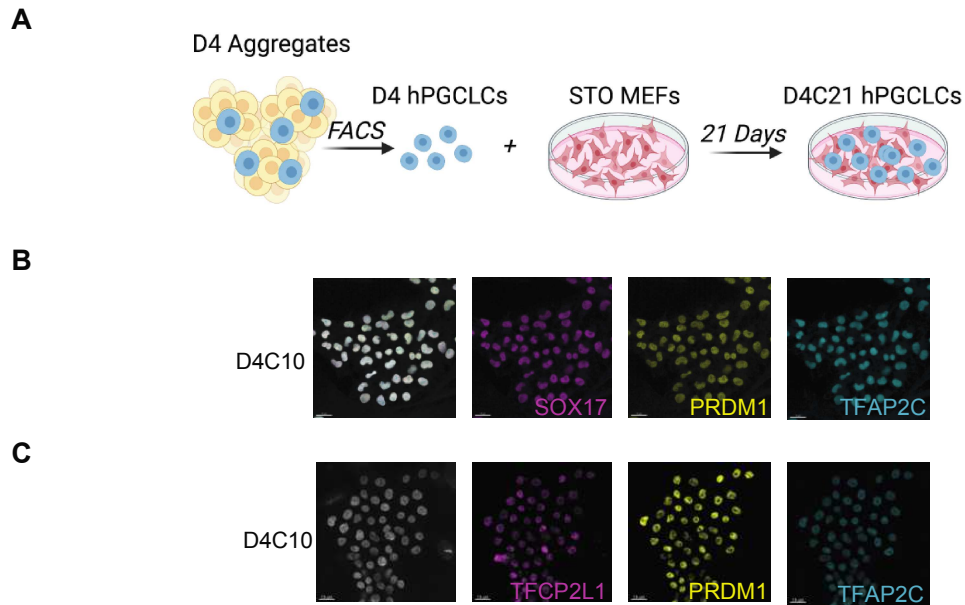


Figure 4. hPGCLCs survive in extended culture and express TFCP2L1

a. Schematic of D4 hPGCLCs grown in extended culture (C). Analysis of extended culture is referred to as D4CX. b. Representative immunofluorescence of UCLA2 hPGCLCs at D4C10. c. Representative immunofluorescence image of wildtype UCLA2 hPGCLCs at D4C21 (n= 10 colonies analyzed) scale bars = 10 microns.

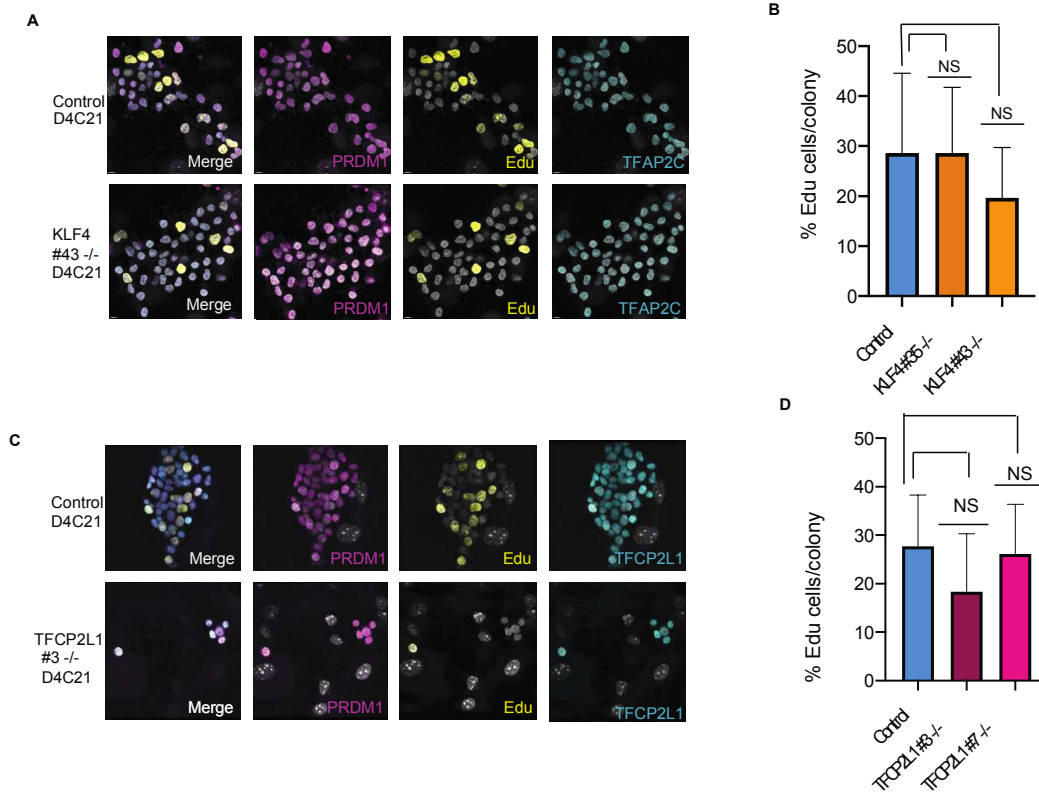
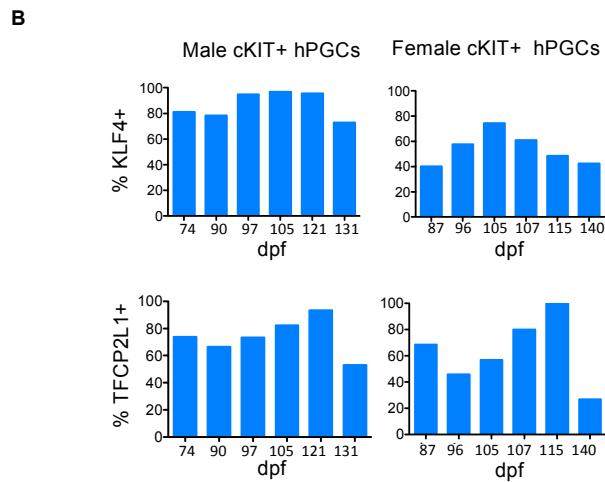
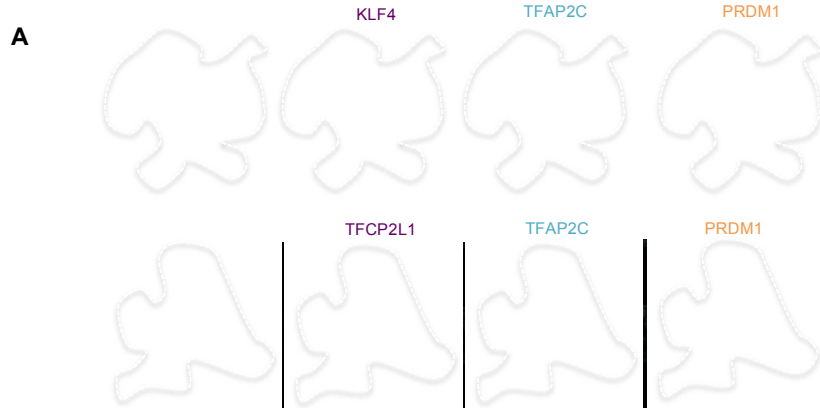
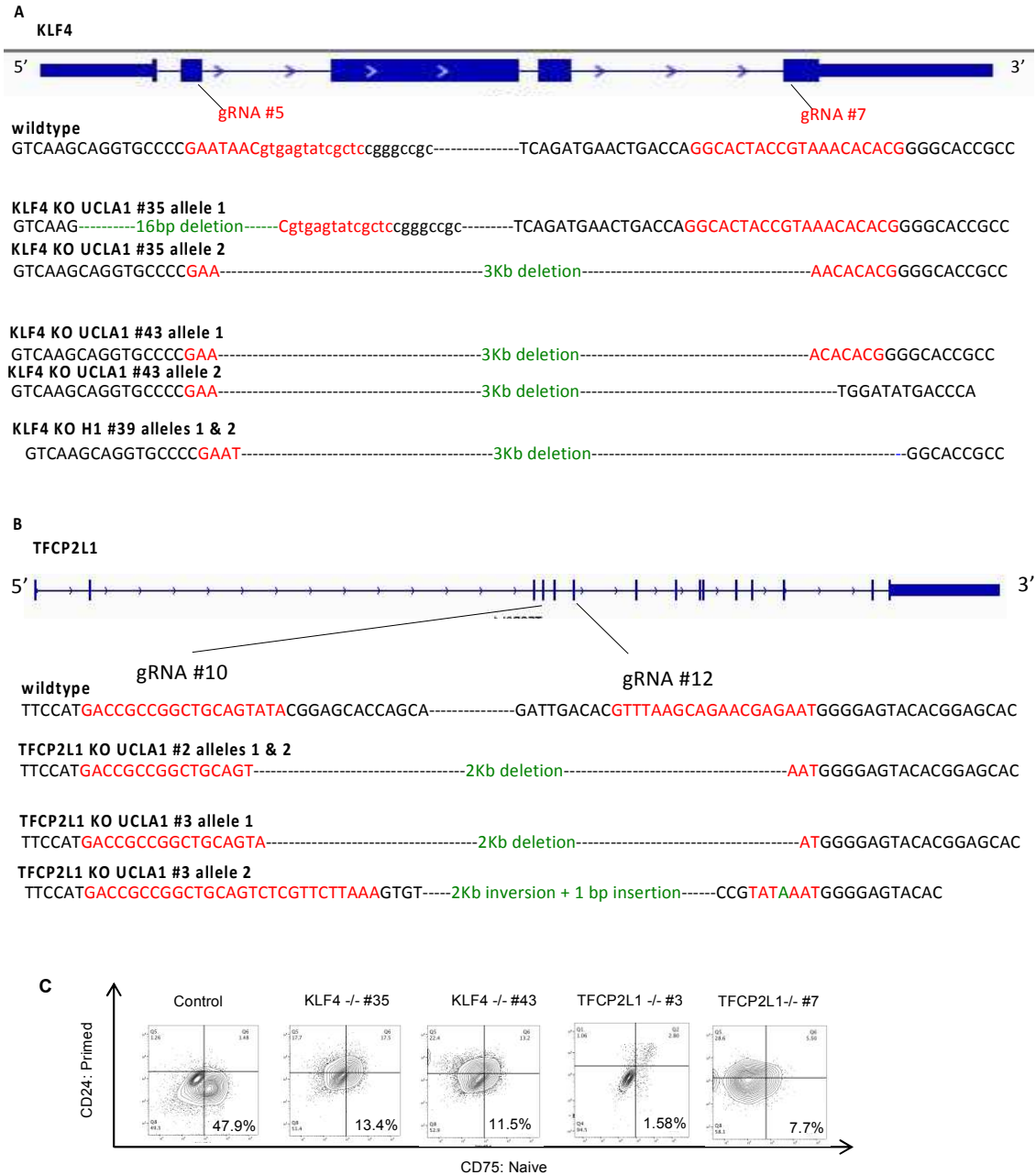


Figure 5. KLF4 and TFCP2L1 are not required for proliferation in extended culture. a. Representative immunofluorescence image of KLF4 control and KO hPGCLCs with Edu staining. b. Quantification of percentage of Edu+ cells within a hPGCLC colony. Error bars indicate n=4-10 colonies analyzed. Scale bars = 10 microns. c. Representative immunofluorescence image of TFCP2L1 control and KO hPGCLCs with Edu staining. d. Quantification of percentage of Edu+ cells within a hPGCLC colony. Error bars indicate n=4-10 colonies analyzed. Scale bars = 10 microns

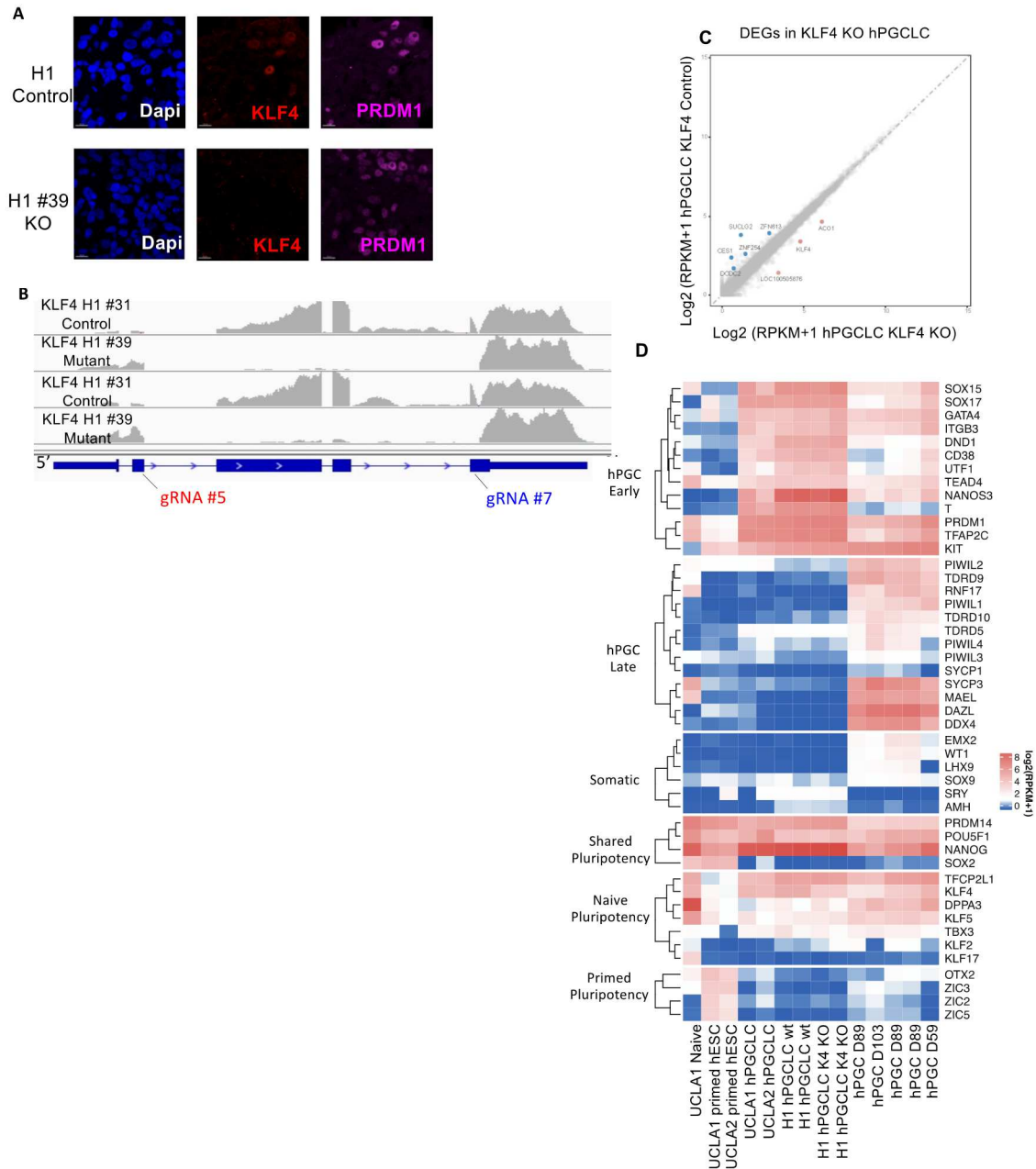


Supplementary Figure 1. KLF4 and TFCP2L1 are dynamically regulated in the germline.

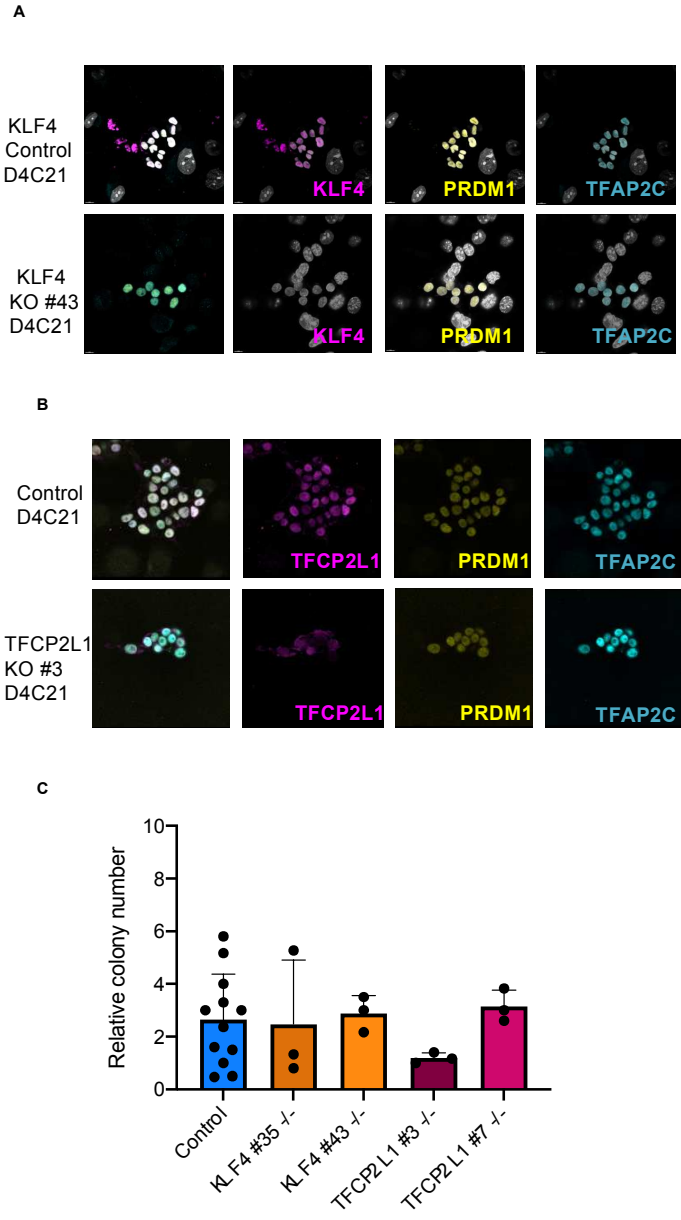
a. Representative immunofluorescence image of KLF4 and TFCP2L1 in TFAP2C/PRDM1 double positive hPGCLCs at D4 of aggregate differentiation using the UCLA2 hESC lines (n= 6-12 independent aggregates analyzed). b. Quantification of the percentage KLF4 or TFCP2L1 positive cKIT⁺ hPGCs at each time as point days post fertilization (dpf). (n= 89-530 cKIT⁺ hPGCs counted per time point).



Supplementary Figure 2. KLF4 and TFCP2L1 mutant hESC subline characterization. a. Schematic of *KLF4* locus for each of the *KLF4* null mutant lines visualized against *KLF4* locus. Red text indicates guide RNA sequence and green indicates resulting mutation. b. Schematic of *TFCP2L1* mutation in each allele for the subclones UCLA1 #7 and #3. c. Flow cytometry analysis for CD75-positive, CD24-negative naive populations at passage 3 (D17-D20) of reversion in *KLF4* and *TFCP2L1* control and KO. Gating set by unstained controls.



Supplementary Figure 3. KLF4 knockout does not affect the hPGC or pluripotency transcriptome. a. Representative image of PRDM1 expressing hPGCLCs in KLF4 control and mutant aggregates at D4 using the H1 hESC sublines (n= 5 aggregates). Scale=15 microns. b. Screenshot of the *KLF4* locus and resulting transcript signal by RNA sequencing in the H1 #31 control and H1 #39 mutant lines. c. RNA-Seq gene expression normalized between KLF4 mutant and control lines highlighting differentially upregulated genes (n=2 independent replicates) d. Heatmap of diagnostic gene expression comparing KLF4 KO and control sublines to hPGCs, hPGCLCs, naive and primed cells. Scale bars = 15 microns.



Supplementary Figure 4. KLF4 and TFCP2L1 do not regulate hPGCLC survival. a. Representative immunofluorescence image of KLF4 control and KO hPGCLCs at D4C21. b. Representative immunofluorescence image of TFCP2L1 control and KO hPGCLCs at D4C21. c. Quantification of number of hPGCLC colonies in each experimental condition normalized to cells plated. Error bars indicate n=3 biological replicates analyzed. Scale bars = 10 microns.

Supplementary Tables

Supplementary Table 1: Guide RNA and primer sequences for CRISPR mutations

	forward	reverse
KLF4 gRNA5	5' GAATAACgtgagtatcgctc	n/a
KLF4 gRNA7	5' CGTGTGTTTACGGTAGTGCC	n/a
KLF4 external of cut genotyping	5' ATTAATGAGgtaggtgaggcg	5'TTCCCTCATCGGGAAGACAGT
KLF4 internal 5' cut genotyping	Same as external forward	5' caccacaccacgaaaacc
KLF4 internal 3' cut	5' tgatggcaaaattgctggga	Same as external reverse
TFCP2L1 gRNA10	5' GACCGCCGGCTGCAGTATA	n/a
TFCP2L1 gRNA12	5' GTTTAAGCAGAACGAGAAT	n/a
TFCP2L1 external of cut genotyping	5' gcactgagggtgtgagaaggct	5' gtctctgtggcgaaatgtgaccac
TFCP2L1 internal 5' cut genotyping	Same as internal forward	5' ccaggagactcacCGATGTCC
TFCP2L1 internal 3' cut genotyping	5' agtgggtgggacttagctggt	Same as external reverse

Supplementary Table 2: RNA sequencing samples and mapping

Sample name	adaptor	Adaptor sequence	Raw reads	# mapped reads	Mapped %
hPGCLC H1-O4-GFP KLF4 control rep 1	BC9	ACACGA	34858644	29095098	83.47
hPGCLC H1-O4-GFP KLF4 KO rep 1	BC10	CACACA	45144724	37308983	82.64
hPGCLC H1-O4-GFP KLF4 control rep 2	BC13	GTGGCC	42998208	35615371	82.83
hPGCLC H1-O4-GFP KLF4 KO rep 2	BC14	TGTTGC	48626147	39620501	81.48

Chapter 6

Conclusion and future perspectives

Given the number of abnormalities that can occur during embryonic and PGC development it is critical that we understand the tightly controlled mechanisms behind these processes. Before implantation of the embryo, the epiblast cells within the blastocyst are at the ground-state, or possess naive pluripotency (Hackett and Surani, 2014; Nichols and Smith; 2017). hPGCs, which are specified after implantation, possess overlapping characteristics with this naive state, but given practical and ethical limitations in human tissue research, little is known about newly specified hPGCs. In Chapter 2 we discussed peri-implantation embryo development and hPGC specification, and described shared and divergent characteristics between mammalian model organisms including the mouse, nonhuman primate species, and pig.

Given these limitations, in this work we relied on *in vitro* stem cell models to study hPGC development, as discussed in the remaining chapters. In chapter 3 we discussed how we optimized CRISPR/Cas9 gene editing to create null mutants of a gene of interest in hESCs before differentiating them to hPGCLC aggregates. Here we demonstrated that one transcription factor, EOMES, is required for hPGCLC induction. Not only will this identified role aid in genetic counseling should a patient with this mutation seek fertility treatment, but this finding also validates a robust system for identifying more genes important for hPGC development. Furthermore, in Chapter 4 we discussed a method of extended culture for hPGCLCs that have begun epigenetic reprogramming, which can be used to evaluate the role of certain genes in later-stage development and proliferation. Finally, in chapter 5 we used these tools to evaluate the roles of KLF4 and TFCP2L1 in hPGC development. Here I discovered that while these transcription factors are required for the re-acquisition of naive pluripotency in hESCs, they do not serve the same role in hPGCLCs.

Pluripotency genes are commonly associated with cancer (Liu et al., 2013; Müller et al., 2016), as is consistent in the germline where mutations including overexpression of NANOG, a copy number mutation of PRDM14, and SNPs in the PRDM14, DAZL, and TFCP2L1 loci are associated with germ cell tumors (Hart et al., 2005; Ruark et al., 2013; Wang et al., 2017). One possibility discussed in Chapter 5 is that the unique regulation of this pluripotency serves a protective role against germ cell tumors, as hPGCs rapidly proliferate during migration. Given the delayed emergence of *TFCP2L1* protein in extended culture and late-stage hPGCs, it is likely that a functional role, protective or otherwise may occur later in hPGC development rather than at specification or in extended culture. To test this we look forward to advancing models where we can study later stage hPGC development. Others in the field are developing embryoid models (Zheng et al., 2019) and reconstituted gonads which allow integrated PGCs to eventually express VASA (Yamashiro et al., 2020). These models have their benefits but also limitations including transient time in culture and inconsistency. Nevertheless, advancing human embryo and gastruloid models will provide more sophisticated insights into the spatiotemporal development of the embryo (Moris et al., 2014; Shahbazi et al., 2016; Warmflash et al., 2014; Zheng et al., 2019) while xenogenic reconstituted ovaries and testes (xrOVaries and xrTestes) provide the potential to study the roles of new genes in later-stage PGC development (Hwang et al., 2020; Yamashiro et al., 2020). Additionally, we share important commonalities with nonhuman primates and will continue to use these for study (Sasaki et al., 2016; Sosa et al., 2018) bearing in mind ethical considerations.

Altogether, this work contributes to our understanding of human development so that we can better study and treat diseases including cancer, infertility, and developmental disorders. Elucidating the genes and mechanisms associated with different stages of hPGC development

will aid in genetic counseling, not only for those experiencing infertility but also to screen for potential developmental and germ cell disorders so that couples can make more informed decisions about their reproductive health. Additionally, a fuller understanding will help us consider preventative measures and targeted therapeutic intervention for other diseases. By using stem cells to model and understand the pluripotent state of hPGC development we now better understand the discrete nature of various pluripotent states and the factors that control them in the finely-tuned processes of human development.

References

- Hackett, J.A., and Surani, M.A. (2014). Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell* *15*, 416–430.
- Hart, A.H., Hartley, L., Parker, K., Ibrahim, M., Looijenga, L.H.J., Pauchnik, M., Chow, C.W., and Robb, L. (2005). The pluripotency homeobox gene NANOG is expressed in human germ cell tumors. *Cancer* *104*, 2092–2098.
- Hwang, Y.S., Suzuki, S., Seita, Y., Ito, J., Sakata, Y., Aso, H., Sato, K., Hermann, B.P., and Sasaki, K. (2020). Reconstitution of prospermatogonial specification in vitro from human induced pluripotent stem cells. *Nature Communications* *11*, 5656.
- Liu, A., Yu, X., and Liu, S. (2013). Pluripotency transcription factors and cancer stem cells: small genes make a big difference. *Chin J Cancer* *32*, 483–487.
- Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. *Nature* *582*, 410–415.
- Müller, M. (2016). The role of pluripotency factors to drive stemness in gastrointestinal cancer. *Stem Cell Research* *16*, 349–357. Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* *4*, 487–492.
- Ruark, E., Seal, S., McDonald, H., Zhang, F., Elliot, A., Lau, K., Perdeaux, E., Rapley, E., Eeles, R., Peto, J., et al. (2013). Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat Genet* *45*, 686–689.
- Shahbazi, M.N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., Fogarty, N.M.E., Campbell, A., Devito, L.G., Ilic, D., et al. (2016). Self-organization of the human embryo in the absence of maternal tissues. *Nature Cell Biology* *18*, 700–708.
- Sosa, E., Chen, D., Rojas, E.J., Hennebold, J.D., Peters, K.A., Wu, Z., Lam, T.N., Mitchell, J.M., Sukhwani, M., Tailor, R.C., et al. (2018). Differentiation of primate primordial germ cell-like cells following transplantation into the adult gonadal niche. *Nature Communications* *9*, 5339.
- Wang, Z., McGlynn, K.A., Rajpert-De Meyts, E., Bishop, D.T., Chung, C.C., Dalgaard, M.D., Greene, M.H., Gupta, R., Grotmol, T., Haugen, T.B., et al. (2017). Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. *Nat. Genet.* *49*, 1141–1147.
- Warmflash, A., Sorre, B., Etoc, F., Siggia, E.D., and Brivanlou, A.H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nature Methods* *11*, 847–854.

Yamashiro, C., Sasaki, K., Yabuta, Y., Kojima, Y., Nakamura, T., Okamoto, I., Yokobayashi, S., Murase, Y., Ishikura, Y., Shirane, K., et al. (2018). Generation of human oogonia from induced pluripotent stem cells in vitro. *Science* 362, 356–360.

Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M., Gumucio, D.L., et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. *Nature* 573, 421–425.