Specification and Epigenetic Programming of the Human Germline

Walfred W.C. Tang[1, 2], Toshihiro Kobayashi[1, 2], Naoko Irie[1, 2], Sabine Dietmann[1, 3] & M. Azim Surani[1, 2, 4]

[1] Wellcome Trust Cancer Research UK Gurdon Institute, Tennis Court Road, University of Cambridge, Cambridge CB2 1QN, UK.

[2] Department of Physiology, Development and Neuroscience, Downing Street, University of Cambridge, Cambridge, CB2 3EG, UK.

[3] Wellcome Trust-Medical Research Council Stem Cell Institute, Tennis Court Road, University of Cambridge, Cambridge, CB2 1QR, UK.

[4] Correspondence: a.surani@gurdon.cam.ac.uk

Abstract

Primordial germ cells (PGCs), the precursors of sperm and eggs, are established in peri-gastrulation stage-embryos in mammals. Signals from extraembryonic tissues induce a unique gene regulatory network in germline competent cells for PGC specification. This network also initiates comprehensive epigenome resetting, including global DNA demethylation and chromatin reorganization. Mouse germline development has been studied extensively, but the extent to which such knowledge applies to humans was unclear. Here, we review latest advances in human PGC specification and epigenetic reprogramming. Although the overall developmental dynamics of human and mouse germline appear similar, there are critical mechanistic differences in PGC specification, reflecting divergence in the regulation of pluripotency and early development.

(109 words)

Key Points

- Regulation of pluripotency and early postimplantation embryonic development have diverged between human and mouse, which might affect the mechanism of primordial germ cell (PGC) specification.
- Specification of human and mouse PGCs occurs in response to extrinsic signals, including BMP2/4.
- Models of human PGC specification from pluripotent stem cells suggests that human PGCs originate from mesodermal precursors at the posterior epiblast during the onset of gastrulation, while mouse PGCs originate from pre-gastrulation epiblast.
- The gene regulatory network for PGC specification and maintenance in humans and mice has diverged. Notably, SOX17, a key endoderm specifier, is also critical for PGC specification in humans but not in mice.
- PGCs undergo genome-wide DNA demethylation which erases parental epigenetic memories and facilitates germ cell differentiation in humans and mice.
- Repressive histone modifications might safeguard PGC genome stability during global DNA demethylation.

Introduction

The germline provides the enduring link between all generations of an organism. In mammals, primordial germ cells (PGCs), the founder cells of the germline, are specified during early embryonic development. Thereafter, PGCs migrate to the developing gonads, where they undergo meiosis and differentiate into gametes. Fertilization of the oocyte by the sperm results in a totipotent zygote that gives rise to all lineages of an organism, including the germline itself. Thus, specification of PGC can be regarded as a crucial first step for the acquisition of totipotency and the continuation of the mammalian life cycle (Figure 1).

Mouse is the primary mammalian model for investigations on germ cell development. Mouse primordial germ cell (mPGC) specification is initiated in the early postimplantation embryo that develops as an egg cylinder (Box 1). Bone morphogenetic protein (BMP) and WNT signals[1] from extraembryonic tissues induce expression of PRDM1 (also known as BLIMP1), the key regulator of PGC fate, in a few pluripotent epiblast cells at ~embryonic day (E) 6.25[2]. This is followed by upregulation of two other key specification factors: PRDM14[3] and TFAP2C[4]. Thereafter, PRDM1, PRDM14 and TFAP2C form a core regulatory network that induces the germ cell fate[5-7]. At E7.25, a founder population of ~40 mPGCs are formed at the base of the allantois[8], which subsequently migrate into the hindgut and colonize the genital ridge by E11.5. To erase somatic epigenetic memories, migratory mPGCs undergo genome-wide epigenetic reprogramming, which entails global DNA demethylation[9-11], imprint erasure[12, 13], X-chromosome reactivation[14-17] and reorganisation of chromatin modifications[18, 19]. These epigenetic changes are attained by ~E13.5, when XY and XX germ cells undergo mitotic arrest and meiotic entry, respectively. This marks the end of the PGC stage of germline development[20].

While the origin of human primordial germ cells (hPGCs) is less clear than that of mPGCs, hPGC specification likely occurs between week 2-3 of development when direct studies on

these early postimplantation human embryos are impractical[21, 22]. Differences in PGC specification mechanisms might be envisaged since regulation of pluripotency and early postimplantation development in mouse and human differ (see Box 1 for details). Specified hPGCs are first observed during the early fourth week (~E24)[23, 24] in the extraembryonic yolk sac wall near the allantois, which is an equivalent location to that of mPGCs at ~E8. Thereafter, hPGCs migrate and colonize the incipient genital ridge by the early sixth week (~E37)[24]. Gonadal hPGCs (also known as gonocytes (XY) or oogonia (XX)) remain proliferative until around the tenth week, when they asynchronously enter mitotic quiescence in male embryos and meiotic prophase in female embryos.

Until recently, our understanding of human germline development was mainly based on extrapolation from mice, supplemented with sporadic immunohistochemistry studies in humans. However, advances in mouse germline biology, particularly over the past 15 years, have informed our more recent breakthroughs on the mechanism of hPGC development. Here, we review the signalling principle and gene regulatory network for human and mouse PGC specification. Emerging evidence shows that there are critical mechanistic differences between human and mouse PGC specification that were not anticipated from studies in mice. We also review the latest advances in germline epigenetic reprogramming in the two species, with particular focus on DNA demethylation and chromatin reorganization. These events have impacts on subsequent germline development and the totipotent state at fertilization, as well as on the transmission of genetic and epigenetic information to subsequent generations.

Mammalian germline induction

Induction of germ cell fate by extrinsic signals

There are two basic modes for germline formation in metazoans: “preformation” and “epigenesis”[25]. In *C.elegans, Drosophila* and *Xenopus*, maternally inherited determinants (known as germ plasm) are exclusively allocated to prospective germ cells following fertilization (i.e. preformation)[26]. However, in mammals and many other animals, germ cells are induced in competent cells by extrinsic signals during embryogenesis (i.e. epigenesis)[1, 27, 28].

In mice, PGC specification occurs in response to complex signalling crosstalk within the egg-cylinder just before gastrulation[1, 29, 30] (Figure 2a). BMP signals produced in the extraembryonic ectoderm (ExE) and proximal visceral endoderm (VE) are essential for mPGC induction in the posterior proximal epiblast at E6.25. Genetic ablation of *Bmp2/4/8b*[31-33] or downstream signalling components (e.g. *Alk2*[34] and *Smad1//4/5*[35-37]) in mice results in loss or reduction of mPGCs. In a comprehensive signalling study, Ohinata *et al*[2] found that both BMP2 and BMP4 can dose-dependently induce PRDM1-postive mPGCs in isolated epiblasts cultured *ex vivo*, with BMP4 being the more potent inducer. In addition, the authors reported that mutants of *Wnt3* and its downstream signalling transducer *β-Catenin,* fail to form mPGCs regardless of BMP signalling[1, 30]. WNT3 is expressed in the posterior VE and epiblast at E5.5-E6.5 and is known to induce primitive streak genes essential for gastrulation[38]. While these mesodermal genes are initially induced in mPGC precursors, they become repressed in specified E7.5 mPGC to maintain germ cell identity. However, one of the WNT3 target genes, mesodermal factor *T* (or *Brachyury*), is unexpectedly required for mPGC specification[30]. Whilst PRDM1-positive mPGC precursors are formed in *T* knockout embryos, they fail to sustain PRDM1 expression and do not upregulate the mPGC specifier PRMD14. In contrast, overexpression of T alone in WNT3 mutants without cytokines is sufficient to induce PRDM1 and PRDM14 expression by activating their putative enhancers. Thus, some mesodermal factors induced by WNT3/β-CATENIN signalling during gastrulation indeed promote germ cell fate. How precisely BMPs and WNT3 cooperate to induce both germ cells and mesoderm at the posterior epiblast is

still unclear. It is possible that the high BMP dosage at the most proximal posterior epiblast favours germ cell induction over mesoderm formation (Figure 2a). Although the role of nodal signalling has not been investigated, it is likely critical for mPGC specification as *Bmp4* and *Wnt3* are absent in *Nodal*-deficient embryos[39].

The signalling mechanism for PGC specification and gastrulation in non-rodent mammals, which develop as a planar embryonic disc instead of an egg cylinder as in rodents, is less clear. To extrapolate patterning in mice to other mammals, some embryologists proposed a flattened model which unfolds the egg cylinder into a planar embryonic disc[40, 41] (see Figure 2b for details). The predicted signalling pattern in this model is roughly in line with patterns recently observed in rabbit peri-gastrulation planar embryos[42, 43]. In rabbits, both *BMP2* and *BMP4* are expressed as a ring-like domain in the peripheral epiblast and trophoblast, as well as in the nascent mesoderm[42]. In contrast, BMP4 and WNT3 antagonists, *CER1* and *DKK1* respectively, are expressed in the anterior hypoblast[43]. These signals likely restrict rabbit PGC induction and gastrulation to the posterior epiblast where *PRDM1*-positive putative PGCs and the primitive streak are observed[42]. There is currently little information on WNT3 and BMP8a expression in non-rodent mammals. Notably, similar BMP signalling pattern is also observed in pig embryos around the time of PGC induction[44]. Overall, despite differences in development, comparative biology suggests that BMP signalling is very likely a conserved signal for PGC specification in mammals, including humans[45-47] (see below).

Germline competence and *in vitro* PGC induction

In mouse embryos, only a few cells in the posterior proximal epiblast acquire the PGC fate while the rest of the epiblast cells differentiate into the ectoderm, endoderm and mesoderm. However, *ex vivo* culture suggests that most of the pregastrulation E5.5-E6.25 epiblast cells are germline

competent[1, 48]. Initial efforts to reconstitute PGC specification *in vitro* resulted in limited success, mainly due to the lack of germline competent pluripotent states (Box 2). Taking *in vivo* development into account, Hayashi *et al*[49] developed a two-step model to reconstitute mPGC specification *in vitro* (Figure 3a). Accordingly, preimplantation epiblast-like naïve mouse embryonic stem cells (mESCs)[50, 51] are first differentiated into postimplantation epiblast-like cells (mEpiLCs)[49]. Upon exposure to BMP4 and defined supporting cytokines, mEpiLCs efficiently give rise to mPGC-like cells (mPGCLCs) that resemble *in vivo* mPGCs. Remarkably, these cells undergo gametogenesis and yield functional sperm and eggs after transplantation into gonads of neonatal/adult mice[49, 52]. This *in vitro* mPGCLC specification system has since been used to address mechanistic aspects of mPGC development (see later).

Until recently, a robust *in vitro* PGC induction model was not available for humans. Human embryonic stem cells (hESCs) grown under conventional condition (Box 2) are believed to be in a "primed" pluripotent state, and exhibit low germline competence[45, 53]. However, a number of recent studies have proclaimed establishment of naive hESC culture conditions (for review, see[54, 55]). These studies also led to a distinct hESC type which self-renew under a four inhibitors (4i) condition (consists of TGFβ , FGF2 and leukemia inhibitory factor (LIF), together with inhibitors of ERK1/2, GSK3β, JNK and p38)[56]. Remarkably, hESCs cultured in 4i condition directly (without any pre-treatment) respond to BMP2/4, and undergo hPGC-like cell (hPGCLC) specification with high efficiency after 4-5 days (up to ~50%) (Figure 3b), as opposed to <5% efficiency with conventional hESCs[45]. These hPGCLCs exhibit a transcription profile that is globally similar to *in vivo* week 7 hPGCs. Notably, hPGCLCs and hPGCs share a unique gene expression profile, including PGC genes (e.g. *PRDM1*, *TFAP2C* and *NANOS3*), pluripotency genes (e.g. *NANOG*, *POU5F1* (or *OCT4*), *TFCP2L1* and *KLF4*), lineage specifiers (e.g. *SOX17*, *GATA4* and *TEAD4*) and cell surface markers (e.g. *CD38* and *ALPL*)[45]. These hPGCLCs likely

represent pre-migratory hPGCs, which have just initiated global epigenetic reprogramming, including global DNA demethylation[45] and imprint erasure[57].

Several recent studies shows that hESC cultured in 4i condition exhibit a transcriptome that is more similar to conventional primed hESCs than to the naïve pluripotent preimplantation epiblast cells in the blastocyst[58-60]. The key distinction between hESC cultured in 4i and conventional conditions is the upregulation of early mesodermal genes (e.g. *T*, *MIXL1*, *FOXA2*, *RUNX1* and *PDGFRA*) in the former, which might signify their competence for germ cell fate[45]. Consistently, two subsequent studies also demonstrate that cells bearing early mesodermal characters have enhanced capability for hPGCLC specification[47, 61]. In one of the reports, human induced pluripotent stem cells (hiPSCs) cultured under a preformulated commercial medium are induced into “insipient mesoderm-like cells” (hiMeLCs), which are germline competent and can give rise to PRDM1- and TFAP2C-positive hPGCLCs with up to 40% efficiency when exposed to BMP2/4[47] (Figure 3b). Similar to hESCs cultured in 4i medium, hiMeLCs exhibit expression of early mesodermal genes. In comparison to mouse pluripotent stem cells, hiMeLCs are considered to be at a developmental state in between that of pre-gastrulating mEpiLCs and anterior primitive streak-like mEpiSCs[47, 62] (Box 2). These independent studies suggest that hPGCs may originate at the onset of gastrulation from mesodermal precursors, rather than pre-gastrulation epiblast as in mice (Figure 3). In fact, specification of PGCs from mesoderm is a well-conserved mechanism in many organisms, including the basal insect cricket[28] and the tetrapod ancestor axolotl[27]. Even in mice where mPGCs are induced prior to mesoderm formation, mesodermal factors, like T, play an essential role in regulating expression of key mPGC specifiers[30].

Acquisition of germline competence

How do cells acquire competence for germline fate? Recent mouse studies reveal dynamic epigenetic changes during the transition from naïve mESCs to germline competent mEpiLCs[63-66]. In particular, regulatory elements of postimplantation epiblast genes accumulate active enhancer marks H3K4me1 and H3K27ac in mEpiLCs while naïve pluripotency-associated enhancers lose their active states[63]. Promoters of developmental genes and germ cell specifiers (e.g. PRDM1 and PRDM14) become bivalently marked by H3K4me3 and H3K27me3, and are considered poised for activation upon mPGC specification[64]. The mechanistic link between germline competence and the rewired enhancer and promoter epigenetic landscape in mEpiLCs requires further investigations. Unlike mEpiLCs and hiMeLCs, hESCs cultured in 4i condition can self-renew while retaining germline competence seemingly indefinitely. This state is also reversible as it can be gained or lost by switching between 4i and conventional conditions[45]. These properties make them tractable models to elucidate the mechanistic basis of germline competence in humans.

Regulatory Network for mPGC specification

A tripartite transcription factor network

The signals that induce mPGC specification also engender gastrulation and mesoderm formation. This accounts for the expression of early mesoderm genes, such as *T*, *Eomes*, *Hoxa1* and *Hoxb1*, in PRDM1-positive mPGC precursors[67, 68]. Although T initially facilitates mPGC specification[30], somatic genes, in general, need to be repressed in nascent mPGCs in order to maintain germ cell identity. Establishment of mouse germ cell fate is governed primarily by PRDM1, PRDM14 and TFAP2C, which are upregulated in mPGC precursors at E6.25-6.75[2-4] (Figure 4a). Knockout of any one of *Prdm1*, *Prdm14* and *Tfap2c* in mice impairs mPGC

specification[2-4, 69]. These embryos exhibit reduced numbers of AP-positive mutant mPGCs, which are lost between E8.5 to E12.5. A common phenotype observed amongst these cells is the de-repression of mesodermal genes[2, 4, 67, 70], suggesting that the three transcription factors together are essential for suppressing the somatic programme. PRDM1 and PRDM14 are also responsible for upregulation of germ cell and pluripotency genes, and initiation of epigenetic reprogramming. The latter is evident from the repression of EHMT1 which promotes H3K9me2 erasure, and other factors that promote DNA methylation, including *Dnmt3a*, *Dnmt3b* and *Uhrf1* (see later)[2, 3, 67]. Importantly, overexpression of any two of the three transcription factors, or *PRDM14* alone, is sufficient to induce mPGCLCs from competent mEpiLCs[5, 7]. Thus, PRDM1, PRDM14 and TFAP2C form a tripartite transcription factor network that is necessary and sufficient for mPGC induction.

In line with other transcription factors, PRDM1, PRDM14 and TFAP2C are independently involved in diverse developmental processes. For instance, PRDM1 is a well known transcriptional repressor involved in lymphocyte differentiation[71] while PRDM14 maintains pluripotency in ESCs[70, 72, 73]. TFAP2C is expressed in a variety of tissues and plays an essential role in trophectoderm specification[74, 75]. Genome-wide DNA binding maps reveal that PRDM1, PRDM14 and TFAP2C cooperate in an intricate manner to upregulate germ cell and pluripotency genes, repress somatic fate, initiate migration and reset the epigenome[5, 6]. In particular, PRDM1 predominantly binds to promoters[5, 64] whilst PRDM14 is enriched in distal regulatory elements[72]. TFAP2C binds to targets of both PRDM1 and PRDM14 to activate or repress target genes[5, 6]. Since none of these factors demonstrate histone-modifying activities, they likely modulate transcriptional activities by recruiting epigenetic modifiers. PRDM14 was shown to interact with SUZ12, a polycomb repressive complex 2 (PRC2) component, to confer repressive H3K27me3 in mESCs[73]. However, two recent interactome studies show that PRDM14, instead of interacting with PRC2 component, forms a complex with ETO family

protein CBFA2T2 (or MTGR1), which is known to recruit histone deacetylases (HDACs) for gene silencing[76, 77]. Deletion of *Cbfa2t2*[76, 77], or interruption of the interaction between CBFA2T2 and PRDM14[76], mimics the effect of *Prdm14* knockout in pluripotency maintenance and mPGC specification. Similarly, PRDM1 has been reported to interact with histone modifiers, including HDAC2, KDM1A, EHMT1 and PRMT5[78].

Pluripotency factors in mPGC development

Re-expression of pluripotency genes is a hallmark of mPGC development. Apart from POU5F1 which is expressed throughout mPGC induction[79, 80], other pluripotency factors, such as SOX2[67, 81] and DPPA3[68, 82], are transiently repressed in epiblast/mPGC precursors and re-expressed in specified mPGCs at ~E7.25-E7.75. Consistent with the presence of pluripotency factors, mPGCs can readily de-differentiate into pluripotent embryonic germ cells (EGCs) under appropriate culture conditions[83-85].

Due to their requirement for preimplantation development, the involvement of pluripotency factors in mPGC specification has been unclear. Notably, conditional knockout/knockdown of *Pou5f1*[86], *Nanog*[87, 88], *Sox2*[89] and *Sall4*[90] after specification result in loss of mPGCs through apoptosis or proliferation defects, highlighting the importance of pluripotency factors in mPGC maintenance. Making use of the mPGCLC induction model, we recently found that overexpression of NANOG alone in mEpiLCs, but not of POU5F1 or SOX2, can induce germ cell fate[66]. Unexpectedly, NANOG induces mPGCLC by direct activation of *Prdm1* and *Prdm14* enhancers independently of BMP4. On the other hand, loss of *Nanog* impairs the efficiency of mPGCLC induction by BMP4. Since NANOG is expressed in the posterior epiblast where mPGCs are specified from, BMP signalling and NANOG might act cooperatively to induce germ cell fate *in vivo*.

A distinct hPGC specification network

Critical role of endodermal factor SOX17

Immunohistochemistry and recent RNA sequencing studies have shown that hPGCs retain a core set of genes essential for mouse germ cell development[21, 22, 57, 91]. This includes specification genes (e.g. PRDM1 and TFAP2C), germline-specific genes (NANOS3, DND1, DAZL and DDX4) and pluripotency factors (e.g. NANOG, POU5F1 and DPPA3). Nonetheless, careful comparison of mouse and human PGC expression profiles has revealed some key differences. For example, hPGCs lack the core pluripotency gene SOX2[92] but express ICM-associated naïve pluripotency factors TFCP2L1 and KLF4[57]. Furthermore, hPGCs express several lineage specifiers, which are absent in mPGCs, including trophectoderm regulator *TEAD4*[57] and endoderm regulator *SOX17*[93].

Among these differences, the absence of SOX2 and the presence of the endoderm specifier SOX17 are particularly noteworthy, since SOX2 is required for mPGC proliferation[89] whilst SOX17 is dispensable for mPGC specification[94]. Even though both SOX17 and SOX2 belong to the SRY-related HMG-box transcription factor family, it is unlikely that SOX17 is simply a replacement for SOX2 in hPGCs, as these proteins belong to different subfamilies. SOX2, together with SOX1 and SOX3, belongs to the SOXB1 subfamily which has critical roles in pluripotency maintenance and neuroectoderm differentiation[95]. On the other hand, SOX17, together with SOX7 and SOX18, are classified as SOXF subfamily and are essential for endoderm differentiation[96], fetal hematopoiesis[97] and cardiovascular development[98]. Moreover, SOX17 cannot replace SOX2 in pluripotency maintenance and somatic cell reprogramming[99]. In fact, overexpression of SOX17 in mESCs induces exit of pluripotency and endoderm

differentiation[100]. In contrast, the closely related SOXB1 subfamily members, SOX1 and SOX3, can replace SOX2's functions as a regulator of pluripotency and neuron differentiation[95, 99]. It is important to note that all SOXB1 members (i.e. SOX1, SOX2 and SOX3) are absent in hPGCs[57, 92], suggesting that SOXB1-mediated functions are excluded in hPGCs.

Surprisingly, SOX17 is amongst the first transcription factors to be upregulated during the specification of hPGCLC from competent hESCs[45]. Knockout of *SOX17* in competent hESCs abolishes hPGCLC specification and this can be rescued by *SOX17* overexpression. PRDM1 expression is compromised in the absence of *SOX17*[45], indicating that it acts downstream of SOX17. As in mice[2, 69], loss of *PRDM1* during hPGCLC specification results in downregulation of some hPGC genes, and de-repression of developmental genes[45, 47, 57]. Expression of key transcription factors, such as SOX17, TFAP2C and POU5F1, are mildly affected. Interestingly, the loss of *PRDM1* in humans causes upregulation of endoderm genes (e.g. *GATA4*, *GATA6* and *FOXA2*)[45], indicating that PRDM1 plays a role in repressing endodermal genes, which could be induced by SOX17[101]. Overall, PRDM1 is important for initiation of the germ cell programme and repression of somatic genes that are presumably induced by upstream BMP signalling and SOX17 (Figure 4b). PRDM1 may also play a role in initiation of global DNA demethylation by repressing the *de novo* methyltransferase *DNMT3B*[47, 57].

Potential partnership of SOX17 and POU5F1

How can SOX17, a transcription factor with diverse roles in development of somatic lineages, be involved in establishment of human germ cell fate? SOX proteins modulate transcriptional activities with a partner transcription factor which binds to an adjacent DNA sequence[98]. It is likely that SOX17 regulates transcription at distinct sets of loci by interacting with different partners in different cellular contexts. Pluripotency factor POU5F1 is one of the known

interaction partners of SOX17[102]. In both human and mouse ESCs, it is well established that POU5F1 acts in conjunction with SOX2 to maintain pluripotency[95]. Interestingly, POU5F1 "switches" partners from SOX2 to SOX17 during primitive endoderm (PE) differentiation from mESCs[103]. As a result, SOX17-POU5F1 binds to enhancers of endoderm genes and apparently activates their expression for PE fate. Notably, one of the targets of SOX17-POU5F1 in mice is *Prdm1*[103]. Since POU5F1 is highly expressed throughout hPGCLC induction from competent hESCs[45], it is possible that similar switching of partners might occur for establishment of human germ cell fate. A possible scenario is that both germ cell genes and endodermal genes are poised for SOX17-POU5F1 activation in human germline competent cells. One of the potential downstream targets, PRDM1, may then act in tandem to activate germ cell genes and repress endodermal genes and other somatic genes. If so, the rapid downregulation of SOX2 upon hPGCLC specification and its continual absence in hPGCs might be necessary to allow SOX17-POU5F1 interaction and avoid neuroectoderm differentiation. Correct dosage of SOX17 and POU5F1 may also be important for hPGC fate, as both transcription factors exhibit dose-dependent action in directing differentiation[100, 104]. Intriguingly, while PRDM1 represses endodermal factors in hPGCs, it is also expressed in PE where the endodermal genes are expressed, indicating that the repressive role of PRDM1 is context-dependent. Another SOX family protein SOX15 is highly expressed in both human and mouse PGCs[47, 91]. SOX15 is the sole member of SOXG group and is co-expressed in mESCs with SOX2[105]. SOX15 physically interacts with POU5F1[105] and shares some common targets with SOX2 in mESCs[106]. However, the loss of *Sox15* in mESCs does not affect maintenance of pluripotency. In addition to SOX family proteins, trophectoderm specifiers TFAP2C and TEAD4, pluripotency factors NANOG, TFCP2L1 and KLF4, all show early expression in hPGCLCs[45, 91]. Their potential functions and combinatorial effects in hPGCLC specification merit further investigations.

The paradox of PRDM14 in hPGC development

Since PRDM14 is indispensable and sufficient for PGC specification in mice[3, 7, 70], its role in hPGC is of particular interest. PRDM14 is strongly expressed in germline competent hESCs but is rapidly downregulated during hPGCLC induction by BMP2/4[45]. Thereafter, *PRDM14* becomes modestly expressed in specified hPGCLCs. As PRDM14 plays an integral role in the core pluripotency circuit in hESCs[107, 108], its rapid suppression is probably necessary for exit of pluripotency. Indeed, overexpression of PRMD14 in hESCs upon random differentiation prevents the upregulation of genes of the three germ layers (e.g. *T*, *MIXL1*, *GATA4* and *PAX6*)[108]. Partial knockdown experiments suggest that PRDM14 may not be required for hPGCLC specification[61], although additional evidence is required to verify this result. In mice, PRDM14 also facilitates global DNA demethylation by repressing *de novo* methyltransferases *Dnmt3a* and *Dnmt3b*[70], but its role in hPGC reprogramming is unclear. Gonadal hPGCs undergoing global DNA demethylation exhibit low and heterogeneous expression of PRDM14[91] with predominant localization in the cytoplasm[45]. Moreover, even in the presence of nuclear PRDM14, conventional hESCs exhibit very strong expression of DNMT3A/3B and hESC genome remains highly methylated[57]. Thus, cumulative evidences suggest that PRDM14 may have a less prominent role in hPGC development. Intriguingly, when transfected into mouse cells, human PRDM14 can however substitute for mouse PRDM14 to rescue defects of *Prdm14* knockout mESCs[109]. This suggests that the two orthologs are functionally conserved and that the behaviour of human PRDM14 may hinge on the precise interacting partners and cellular contexts. The paradox of PRDM14 in human germline development remains to be fully addressed.

Overall, the critical role of SOX17 in hPGC specification and the apparent diminished role for PRDM14 suggest a clear divergence in the mechanisms for PGC specification and the initiation of epigenetic reprogramming between humans and mice. It will be of interest to establish

whether SOX17-mediated germline induction is a general mechanism conserved amongst non-rodent mammals which develop as planar embryonic disc.

Resetting the germline epigenome

In mammals, global epigenetic reprogramming occurs during preimplantation development, and in the early germline. The former resets the zygotic epigenome for naïve pluripotency (reviewed in[110, 111]), while the latter erases parental epigenetic memories and facilitates gametogenesis. Here, we focus on epigenetic resetting in the mouse and human early germline, and discuss its potential impact on germ cell maintenance, differentiation and epigenetic inheritance.

Genome-wide DNA demethylation

In mice, PGCs are specified from the postimplantation epiblast, which is hypermethylated and primed for lineage differentiation. To reset the epigenome, mPGCs undergo genome-wide DNA demethylation[11], X-chromosome reactivation[14-17] and chromatin modification reorganization[19] as they migrate and colonize the genital ridge from E8.0 to E13.5 (Figure 5). During this time, global CpG methylation levels drop from ~70% in the epiblast to unprecedented low levels of ~4% in E13.5 mPGCs[9, 112]. As a result, almost all genomic features, including imprint control regions, become hypomethylated. Methylation is then re-established in a sex-specific manner after E13.5 in males and after birth in females[110].

The current view is that global DNA demethylation in the germline is primarily achieved through a passive mechanism. Shortly after mPGC specification, PRDM1 and PRDM14 repress expression of *de novo* DNA methyltransferases DNMT3A and DNMT3B, as well as UHRF1[5, 67],

a recruitment factor of DNMT1 that is essential for the maintenance of DNA methylation[113]. As a result, both maintenance and *de novo* methylation activities are apparently repressed, enabling replication-coupled DNA demethylation when mPGCs proliferates[9, 114-116].

Recent evidence suggests that enzymatic conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcyotsine (5hmC) also plays a part in demethylation, especially for imprints, which are protected from demethylation until mPGCs settle in the genital ridge at ~E9.5-E10.5. Ten-eleven translocation enzymes (TET1, TET2 and TET3) oxidize 5mC to 5hmC and downstream derivatives, which can subsequently be enzymatically removed and replaced by unmethylated cytosine, or be diluted passively during replication[117]. Concurrent with TET1 and TET2 upregulation at E9.5-E11.5, transient global increase of 5hmC coupled with reduction of 5mC is observed in mPGCs[10, 118]. Nonetheless, knockout studies in mice suggest that a large extent of germline DNA demethylation can still occur independently of TET1 and TET2[119-121]. However, TET1 and TET2 are required for efficient erasure of imprints[119, 120] and demethylation of meiotic gene promoters for germ cell differentiation[121].

Recently, three independent studies showed that the overall DNA demethylation dynamics in hPGCs is similar to that in the mouse germline (Figure 5a)[57, 91, 122]. At week 5, migratory hPGCs in the hindgut already exhibit low 5mC levels compared to neighbouring soma[57]. As hPGCs settle in the genital ridge, DNA methylation levels drop to a minimum of ~4.5% by week 8[57, 91]. Interestingly, germ cells of both sexes (isolated by hPGC surface marker c-KIT) stay hypomethylated at week 19[91], around 9 weeks after the start of meiotic entry in females and mitotic quiescence in males. Immunohistochemistry studies show that this c-KIT positive population represents rare germ cells that retain hPGC characteristics, while the majority of germ cells have differentiated into c-KIT-negative prespermatogonia or oogonia[123-125]. While human oogonia likely remain hypomethylated before birth as in mice, it remains unclear whether the more advanced male germ cells at week 19 have begun remethylation[126].

As in mPGCs, DNA demethylation in hPGCs is associated with the repression of UHRF1, DNMT3A and DNMT3B. Furthermore, hPGCs exhibit transient high levels of 5hmC, which is coupled with TET1 and TET2 upregulation[57]. Notably, the majority of imprints are already demethylated when hPGCs arrive at the genital ridge[57], indicating earlier imprint erasure dynamics in humans as compared with mice.

Chromatin reorganization safeguards the germline

DNA methylation is an important epigenetic silencer, which modulates gene expression and maintains genome stability in mammalian cells[127]. Loss of DNA methylation in somatic lineages causes derepression of retrotransposons, proliferation defects and cell death. Notably, hypomethylated human and mouse PGCs remain proliferative without signs of global gene or retrotransposon activation. Similar to PGCs, *Dnmt1*/*Dnmt3A*/*Dnmt3B* triple knockout mESCs can self-renew in the absence DNA methylation[128], although they perish upon differentiation. Recent studies in mESCs suggest that global DNA demethylation triggers reorganization of repressive chromatin modifications to repress retrotransposons and maintain genome stability[129, 130]. In fact, DNA demethylation in mPGCs is also accompanied by global chromatin modification reorganization[18, 19]. Apart from depletion of H3K9me2, repressive H3K27me3 and H2A/H4R3me2s are both enriched during the course of mPGC development, while H3K9me3 is retained predominantly at pericentric heterochromatin (Figure 5a)[19, 131]. In addition to modulating gene expression, these marks have been implicated in the repression of retrotransposons in mPGCs[132]. For instance, loss of H3K9 methyltransferase SETDB1 in mPGCs causes depletion of H3K9me3 and H3K27me3 in endogenous retroviruses (ERVs)[133]. This is associated with depression of many ERVs (e.g. intracisternal A particle (IAP)) and reduced numbers of male germ cells. Likewise, loss of arginine methyltransferase PRMT5 in mPGCs results in male and

female sterility that is associated with depletion of H2A/H4R3me2s and upregulation of LINE1 and IAP elements in mPGCs[134]. Global reorganization of repressive histone modifications is also observed in hPGCs, albeit with slightly different dynamics (Figure 5a). Thus, it is likely that one of the purposes of germline chromatin reorganization is to safeguard genome integrity while PGCs undergo DNA demethylation to unprecedentedly low levels (Figure 5b and 5c).

Notably, naïve pluripotency genes (e.g. TFCP2L1 and KLF4) are expressed in human ICM, naïve hESCs and hPGCs, all of which exhibit a globally hypomethylated genome[57, 60, 135, 136]. It is tempting to speculate that naïve pluripotency factors may have roles in facilitating genome-wide DNA demethylation and/or safeguarding the globally hypomethylated genome. However, the comprehensive extent of DNA demethylation seen in hPGCs (down to 4% average CpG methylation in contrast to 30-40% in ICM and naïve hESCs) suggests that additional factors are in place to facilitate robust DNA demethylation.

“Escapees” from global DNA demethylation

Despite global DNA demethylation, some genomic loci, referred to as “escapees”, remain methylated in human and mouse PGCs[9-11, 57]. In both species, the vast majority of escapees are associated with retrotransposable elements[9-11, 57, 91]. In particular, evolutionarily young and potentially hazardous retrotransposons, such as IAP in mice[137] and SVA in humans[57], remain relatively highly methylated, which may contribute to their repression. Escapees are also found in pericentric satellite repeats[57] and in subtelomeric regions[11]. Retention of DNA methylation in these regions may maintain chromosome stability and ensure proper chromosome alignment and segregation during mitosis. Notably, a minority of escapees are single copy sequences[9-11]. In humans, “repeat-poor” escapees are widely distributed in the genome, including promoters, gene bodies and enhancers[57]. Some escapees are associated with genes that are expressed in

the brain and participate in neuronal differentiation, while some are found in ubiquitously expressed genes, like circadian regulator *CSNK1D*[57]. The functions of these repeat-poor escapees, particularly those found in regulatory regions, remain to be determined.

Potential mechanisms for escapee methylation

The existence of escapees suggests that DNA methylation pathways are not completely repressed in PGCs. In both human and mouse PGCs, DNMT1 remains strongly expressed in the nucleus, whilst DNMT3A, DNMT3B and UHRF1 are not detectable[57, 114, 122]. While DNMT1 is generally regarded as a maintenance DNA methyltransferase, recent studies suggest that DNMT1 exhibits *de novo* methylation activity on unmethylated DNA *in vivo* and *in vitro*[138]. Given these evidences, it is possible that DNMT1 acts independently of UHRF1 to confer and/or maintain methylation at escapee loci. Human escapees are enriched for H3K9me3 and KAP1 binding sites[57], suggesting that DNMT may be targeted to the escapees via the KAP1/KRAB-ZFP co-repressor complex. The zinc finger domain of KRAB-ZFPs binds to specific DNA sequences while the KRAB domain interacts with KAP1, which recruits histone deacetylases, histone methyltransferase SETDB1 and/or DNMT1 for heterochromatin formation[139]. In humans, there are approximately 400 KRAB-ZFPs, presumably with different DNA binding specificity, and hence different targets. For instance, ZFP91 and ZFP93 have been shown to target and repress human SVAs and L1PAs retrotransposons respectively[140], while ZFP57 maintains methylation at imprinted loci[141, 142]. Many KRAB-ZFPs are highly expressed in hPGCs and may be involved in recruiting DNMT1 to confer DNA methylation at retrotransposon and single copy escapee loci[57]. Parallel mechanisms, like PIWI-piRNA pathways, may also take part in methylation of retrotransposons[57, 91, 143]. Notably, PRDM9 (also known as ZNF899) has been

shown to be essential for meiotic recombination[144], suggesting that KRAB-ZFPs may also perform other critical functions during germ cell development.

Epigenetic inheritance via DNA methylation

The inheritance of non-genetic information through the mammalian germline has received considerable attention in recent years[145-147]. Since DNA methylation is stably heritable through cell division, it is a candidate for the transmission of epigenetic information through the germline. Indeed, a classical example of epigenetic inheritance involving DNA methylation is the transmission of genomic imprints[148, 149]. Parent-of-origin-specific DNA methylation is established at imprint control regions (ICR) concurrently with global remethylation during gametogenesis and is transmitted to the zygote following fertilization. Although imprints are maintained during global demethylation in the preimplantation embryo and persist in somatic cells, they are erased during germline reprogramming and re-established in each generation. Imprinting is therefore an intergenerational epigenetic inheritance phenomenon. Indeed, the comprehensive nature of germline and, to a lesser extent, DNA demethylation during preimplantation development would reduce the likelihood of transgenerational epigenetic inheritance (TEI) through DNA methylation[147]. Moreover, a large portion of the genome becomes fully remethylated during gametogenesis, representing another barrier to TEI. However, recent studies have identified some single copy and retrotransposon loci that can escape both waves of demethylation[9, 11, 57] and some of these loci remain partially methylated in gametes[57]. It is possible that methylation at these escapee loci might be susceptible to environmental factors and such epigenetic information could potentially be transmitted to subsequent generations with phenotypic consequence.

Inheritance of environment-induced metabolic and behavioural traits have been reported in mammals, where most phenotypes last for one to two generations (reviewed by[145, 146]). For instance, male mice exposed to *in utero* undernutrition yield offspring that exhibit obesity and glucose intolerance[150]. Although there are locus-specific DNA methylation changes in sperm of exposed males, differential methylation is not detected in the offspring. More recently, it was shown that paternal diet does not have consistent effects on the sperm methylome[151] while small RNA in sperm may contribute to intergenerational inheritance of diet-induced metabolic disorders[152, 153]. Taken together, recent evidence indicates that DNA methylation is less likely to be the primary mechanism for environment-induced epigenetic inheritance, but this cannot be entirely excluded at present. Moreover, it remains possible that some retrotransposon escapees, such as IAP in mice and SVA in humans, may give rise to metastable epialleles, which are vulnerable to environmental influence. This is exemplified by *Agouti viable yellow* mice and *Axin-fused* mice, in which IAP insertion to the vicinity of the *Agouti* or *Axin-fused* genes causes ectopic gene expression and TEI phenotypes[145].

Conclusions and Perspectives

The divergence of human and mouse occurred ~60 million years ago. Since then, mice have acquired a unique egg-cylinder structure for peri-gastrulation development[154], while humans and a number of other mammals develop as a planar embryonic disc. The germline of both species arise during this period of developmental divergence. Recent breakthrough in *in vitro* hPGCLCs specification and direct studies on *in vivo* hPGCs has provided extensive information on human germline development. Whilst BMP signalling appears to be a conserved pathway for human and mouse germ cell induction, the competent state and gene regulatory network for PGC specification have diverged between the two species. On the other hand, the extent and the

dynamics of epigenetic reprogramming in human and mouse PGCs appear similar, albeit not identical. The mechanism of initiation of the epigenetic programme however may differ due to the involvement of SOX17 as a key germline specifier, and as yet unclear contribution of PRDM14. Notably, both germlines feature some single copy and retrotransposon demethylation escapees, which may be important for genome stability and could have potential for epigenetic inheritance.

In the light of these discoveries, new questions emerge that merit further investigation. What is the molecular basis for germline competent states? How do transcription factors with diverse functions in different lineages act combinatorially to specify and maintain germ cell fate? What triggers epigenetic reprogramming in hPGCs? How are “escapees” targeted and do they have functional significance in the germline? Since functional studies in human are not possible, further advances in *in vitro* hPGC development are required to address some of these questions. Under current conditions, hPGCLCs do not progress beyond pre-migratory stage and therefore do not undergo comprehensive epigenetic reprogramming or upregulate meiotic genes[45, 47]. Development of *in vitro* conditions that enable robust differentiation of hPGCLC towards later gonadal stages and beyond will be essential. Indeed, a recent report shows that haploid mouse spermatid-like cells can be derived from mESCs using an entirely *in vitro* step-wise protocol[155]. These haploid cells apparently yield fertile offspring after injection into oocytes, albeit at a low frequency (~2%). While observations in this study remain to be reproduced, it has raised the prospects for *in vitro* reconstitution of meiosis and gametogenesis in humans.

Mice have been and will remain a valuable model for mammalian development. However, studies on human germline development suggest that not all observations in mice can be faithfully extrapolated to humans. Indeed, the mechanism for neuroectoderm specification has also diverged between men and mice[156]. In view of these differences, studies on non-rodent mammals, such as rabbits[42], pigs[44] and non-human primates, in combination with *in vitro* human

models, will provide insights on the mechanism of PGC specification and other early cell fate decision in humans.

Acknowledgement

We thank Jamie Hackett for critical reading of the manuscript and members of the Surani lab for helpful discussions. Our work was funded by Wellcome Trust Investigator Award and BIRAX Initiative to M.A.S. and by a Croucher-Cambridge International Scholarship to W.W.C.T.. The work at the Institute is funded by a core grant from the Wellcome Trust (092096) and Cancer Research UK (C6946/A14492). We apologize to colleagues whose work could not be cited due to length limitations.

Competing Interests statement

The authors declare no competing interests.

Box 1: Divergence of early embryo development in mice and humans

After fertilization, human and mouse zygotes both undergo a series of cleavage divisions (see the figure) and cell fate decisions, leading to the formation of blastocysts (reviewed in[157, 158]). The blastocyst is composed of a pluripotent inner cell mass (ICM) and an outer trophectoderm (TE) (see the figure). The ICM subsequently segregates into preimplantation epiblast and hypoblast (also known as primitive endoderm)**.** The epiblast primarily develops into the embryo proper while the TE and the hypoblast gives rise to extraembryonic tissues, including the placenta and the yolk sac. Whilst human and mouse preimplantation development are morphologically similar, there are notable differences in developmental timing, gene expression, signalling requirement for pluripotency and lineage segregation[58, 159]. For instance, although human and mouse preimplantation epiblasts share the expression of some naïve pluripotency factors (e.g. *TFCP2L1, KLF4* and *TBX3*), *ESRRB* and *KLF2* are present in mice but not in humans[58, 160]. Furthermore, TGFβ signalling pathway components (e.g. *NODAL*, *GDF3* and *TGFBR1*) are highly expressed in the human epiblast[58]. Inhibition of this pathway in blastocysts downregulates the core pluripotency factor NANOG in human epiblast but has no detectable effect in mice, indicating that regulation of pluripotency differs between human and mouse.

Human and mouse embryos become structurally distinct following implantation. As the mouse blastocyst implants at ~E4.5, the pluripotent epiblast cells become polarized and transform into a cup-shaped epithelium[40, 161], while the TE-derived extra-embryonic ectoderm (ExE) forms an inverted cup on top. Concurrently, the hypoblast develops as the visceral endoderm (VE) and envelops the epiblast and ExE. These result in the formation of an egg cylinder with proximal-distal and anterior-posterior axes[157]. At E5.5, the distal VE (DVE) thickens and forms a specialized signalling region, and migrates proximal-anteriorly and become the anterior VE

(AVE). Reciprocal signalling interactions between the epiblast, the ExE and the VE lead to gastrulation at the posterior epiblast at E6.5, where the primitive streak emerges and gives rise to mesoderm and definitive endoderm (reviewed in[29]).

Humans, and possibly all non-rodent mammals, exhibit a planar structure during peri-gastrulation period[162]. Following implantation at E8-9, the epiblast moves away from the trophoblast, creating the amniotic cavity. The postimplantation epiblast and the underlying hypoblast flatten into a round bilaminar embryonic disc, which is sandwiched between the fluid-filled amniotic cavity and the yolk sac (formerly the blastocoel). Cells originated from the epiblast and the hypoblast form the linings of the amniotic cavity (amnion epithelium) and the yolk sac (yolk sac epithelium), respectively. At around E16, gastrulation commences and the primitive streak is formed at the posterior end of the embryonic disc and extends halfway across the epiblast towards the anterior region. The proliferating epiblast cells along the streak undergo epithelial-mesenchymal transition and migrate to the space between the epiblast and PE, giving rise to the mesoderm layer. The ingressing epiblast cells also replace hypoblast cells to become definitive endoderm. This process sees the bilaminar embryonic disc transforming into a trilaminar disc, which contains the three germ layers for lineage specification.

In mice, PGC specification occurs at the posterior epiblast prior to gastrulation[2]. This is dependent on pluripotent state[49] and intricate signalling interaction between embryonic and extraembryonic tissues in the egg cylinder[1, 30] (Figure 2a). Less is known about PGC origin and peri-gastrulation patterning in planar embryos of humans and other non-rodent mammals. The divergence in pluripotency regulation and embryonic structure between rodents and other mammals may culminate in mechanistic differences for germline establishment.

Box 2: Pluripotent states and initial attempts of *in vitro* PGC induction

Pluripotency, a transient state during early embryonic development, can be maintained *in vitro* as self-renewing pluripotent stem cells (for review, see[54, 55, 163]). In mice, naïve pluripotent cells in the ICM of E3.5-E4.5 blastocysts give rise to mouse embryonic stem cells (mESCs) that are traditionally maintained under fetal calf serum and leukemia inhibitory factor (LIF) *in vitro*[164, 165]. Serum/LIF mESCs can contribute to all somatic lineages and the germline when introduced back into blastocysts. Since serum contains various undefined signalling molecules, serum/LIF mESCs are in metastable states and cycle in and out of naïve pluripotency[163]. On the other hand, mouse epiblast stem cells (mEpiSCs) are traditionally derived with Activin A and FGF2 from E5.5-E6.5 mouse postimplantation epiblasts that are poised for lineage differentiation[166, 167]. Although mEpiSCs express core pluripotency factors *Pou5f1*, *Nanog* and *Sox2*, they upregulate somatic markers (e.g. *Otx2*, *Eomes*, *Foxa2* and *T*) and lack expression of naïve pluripotency genes (e.g. *Prdm14*, *Tbx3 and Zfp42* (also known as *Rex1*)). They can contribute to chimaeric embryos when grafted into postimplantation epiblast[62, 168] but exhibit limited contribution when injected into preimplantation blastocysts[166, 167]. Thus, mEpiSCs are considered to be in a “primed” pluripotency state.

Putative germ cells are generated at a very low frequency by spontaneous differentiation of serum/LIF mESCs *in vitro*[169, 170]. The fact that *in vivo* mPGCs are specified from ~E6.0 postimplantation epiblasts led to the speculation that postimplantation-epiblast derived mEpiSCs might be germline competent. However, mEpiSCs also demonstrate very limited capacity for mPGCLC specification (up to 1.5%)[171]. Recent studies show that mEpiSCs acquire properties more similar to anterior primitive streak cells than to pre-gastrulating epiblasts[62], which might explain the low germline competence in these cells. To achieve robust mPGCLC induction, an

intermediate entity in between naive mESCs and primed mEpiSCs is required. Recently, Hayashi et al. established such a germline competent epiblast-like state *in vitro*[49] (Figure 3a).

Human ESCs (hESCs) can also be derived from blastocyst culture[172]. Although hESCs and mESCs have similar embryonic origins, they exhibit different morphology, gene expression patterns and signalling requirements for maintenance of pluripotency[173]. For example, while mESCs are dependent on BMP and LIF signallings, conventional hESCs require FGF and TGFβ signallings[174]. Instead, hESCs share similar properties with mEpiSCs and are considered to be in a "primed" pluripotency state[173]. However, there are some notable differences between hESCs and mEpiSCs. For example, hESCs express naïve pluripotency genes PRDM14 and ZFP42, which are absent in mEpiSCs, and both factors are essential for maintenance of hESC pluripotency[107, 175]. Thus, conventional hESCs likely represent a distinct pluripotent state which has characteristics of both naïve mESCs and primed mEpiSCs.

Previous studies showed that conventional hESCs can differentiate into hPGCLCs at a low frequency (up to 5%) by spontaneous differentiation, which can be promoted by BMP signalling[46, 53, 176]. The low germ cell induction capacity suggests that hESCs do not represent a germ cell competent state. Recent evidence indicates that cells with early mesoderm characteristics exhibit high competence for human germ cell fate[45, 47, 61] (Figure 3b).

Figure legends

Figure 1: Cycle of human germline development

After fertilization, the zygote develops into the blastocyst which contains pluripotent preimplantation epiblast cells. These cells give rise to all lineages in the embryo proper, including the germline. As the blastocyst implants into the uterine wall, it develops a bilaminar

embryonic disc and later undergoes gastrulation to form the ectoderm, mesoderm and endoderm germ layers (see also Box 1). Human primordial germ cells (hPGCs) are likely specified around the time of gastrulation (~E17) although the exact origin of hPGC in the embryo remains unknown. At week 4, hPGCs locate at the yolk sac wall close to the allantois and later migrate through the hindgut to the developing genital ridges. Migratory hPGCs undergo genome-wide epigenetic reprogramming, including global DNA demethylation, to erase imprints and other somatic epigenetic memories. During fetal development and adulthood, gonadal germ cells undergo meiosis and gametogenesis to differentiate into sperm and eggs. Concurrently, the genome is remethylated and acquires appropriate epigenetic signatures for the generation of a totipotent zygote upon fertilization.

Figure 2: Signalling for mammalian germline induction.

a | At the time of mouse primordial germ cell (mPGC) specification and gastrulation, the embryo forms an egg cylinder with anterior-posterior (A-P) and proximal-distal (P-D) axes (left panel). BMP4 from the extraembryonic ectoderm (ExE), together with BMP2 and WNT3 from the posterior visceral endoderm (VE), induce a few PRDM1-positive mPGC precursors at the most posterior proximal epiblast, the site with highest levels of BMP signals[1, 30]. The anterior VE (AVE) expresses antagonists of BMP (e.g. CER1 and LEFTY1) and WNT (e.g. DKK1), which prevents mPGC and mesoderm induction at the anterior epiblast[1]. *Smad2* and *FoxH1* mutant embryos, which lack a functional AVE, exhibit ectopic induction of PRDM1-positive cell in the anterior epiblast[1]. BMP8b expressed in the ExE may have a role in restricting the AVE domain from extending to the posterior epiblast[1]. The expression patterns of key signals for mPGC specification are illustrated on the right panel.

b | A hypothetical signalling model for PGC induction in humans and other non-rodent mammals. The mouse E6.5 embryo is transformed into a planar structure (left panel)[40, 41] which corresponds to ~E17 trilaminar human embryo (right panel). In this model, the cup-shaped mouse epiblast and VE are unfolded into the upper and the lower layer of an embryonic disc respectively, with the nascent mesoderm sitting in between the two layers at the posterior end (left panel). The ExE becomes the structural equivalent of amnion epithelium/peripheral epiblast in humans. The extraembryonic VE in mice would become the yolk sac epithelium. As a result, the proximal-distal axis in the egg cylinder becomes a peripheral-central axis. Projection of key PGC specification signals predicts that BMP4 and BMP8b would be expressed as a ring-like domain that surrounds the epiblast, while the posterior hypoblast/epiblast would be the site of BMP2 and WNT3 signals. Inhibitory signals from the anterior hypoblast would restrict PGC induction to the posterior epiblast/nascent mesoderm, where BMP dosage is the highest. It is important to note that human and mouse postimplantation development are fundamentally different and this model serves to translate signalling patterns in mice for comparison with that in non-rodent mammals.

Figure 3: Reconstitution of mouse and human PGC specification *in vitro*.

a | In mice, the E3.5-E4.5 preimplantation epiblast can give rise to naïve self-renewing embryonic stem cells (mESCs) under a basal culture condition that consists of leukemia inhibitory factor (LIF), GSK3β inhibitor and ERK1/2 inhibitor (referred as 2i/LIF)[50]. Priming of naïve mESCs with Activin A and FGF2 for 2 days yields epiblast-like cells (mEpiLCs), which are transcriptionally similar to E5.75 pre-gastrulating epiblast[49]. Like its *in vivo* counterpart, mEpiLCs are germline competent and can robustly give rise to mouse PGC-like cells (mPGCLCs) in response to high dosage of BMP4 with BMP8B, LIF, stem cell factor (SCF) and epidermal

growth factor (EGF). Transcriptome and epigenetic profiles of day 6 mPGCLCs are globally similar to *in vivo* E9.5 migratory mPGCs. XY mPGCLCs can differentiate into sperm after transplantation into the seminiferous tubules of *W*/*W*^{v} neonatal mice which are sterile and lack endogenous germ cells[49]. The mPGCLC-derived spermatozoa can yield healthy fertile offspring following intracytoplasmic sperm injection into oocytes. Similarly, XX mPGCLCs can mature into functional oocytes following aggregation with female gonadal somatic cells and transplantation into the ovarian bursa of adult mice[52].

b | Despite their origin from blastocyst, conventional human embryonic stem cells (Conv hESCs) cultured under FGF2-containing medium are considered to be in a primed pluripotent state similar to that in postimplantation epiblast and are not germline competent (see also Box 2). Conv hESCs can reversibly adapt to a recently reported "4i" condition, which consists of TGFβ, FGF2 and leukemia inhibitory factor (LIF), together with inhibitors of ERK1/2, GSK3β, JNK and p38[56]. Self-renewing 4i hESCs exhibit expression of early mesoderm genes and are competent for human PGC-like cell (hPGCLC) formation in the presence of BMP2/4 and other cytokines (up to 50% efficiency)[45]. Alternatively, human induced pluripotent stem cells (hIPSCs) cultured under feeder-free condition with an preformulated commercial medium can give rise to hPGCLCs at ~20% efficiency in response to BMP4[47]. When cultured under Activin A and GSK3β inhibitor for 2 days, these hIPSCs can differentiate into incipient mesoderm-like cells (hiMeLCs) which exhibit enhanced competence for hPGCLC formation (up to 40% efficiency). hPGCLCs generated from 4i hESCs and hiMeLCs share similar transcriptome and likely represent pre-migratory stage hPGCs[47, 57].

Figure 4: Gene regulatory network models for mouse and human PGC specification.

a | In mice, BMP-SMAD and WNT3-β-CATENIN signalling induces expression of PRDM1 and PRDM14, as well as mesoderm genes in mPGC precursors[30]. The precise mechanism for initial upregulation of PRDM1 at E6.25 remains unclear, but it is known that mesodermal factor T is required for sustaining PRDM1 expression and triggering PRDM14 expression[30]. Maintenance of PRDM1 and PRDM14 expression are mutually interdependent[3, 70]. PRDM1 and PRDM14 induces TFAP2C expression, and together, the three transcription factors form a core specification network (yellow eclipse)[5, 7] to upregulate germ cell and pluripotency genes, repress mesoderm genes, temporarily inhibit cell proliferation[18], and initiate epigenetic reprogramming and migration.

b | In humans, SOX17 is critical for hPGC specification and is upstream of PRDM1 expression[45]. PRDM1 represses SOX17-induced endoderm genes, BMP/WNT-induced mesoderm genes and other somatic genes. Together, SOX17 and PRDM1 establish the human germ cell programme, while the role of TFAP2C and PRDM14 in hPGC specification remains to be clarified. Notably, SOX2 is absent in hPGCs. BMP signalling[177] and PRDM1[178] may contribute to rapid downregulation of SOX2 during hPGC specification. It is unclear whether BMP-SMAD signalling directly activates *SOX17* or indirectly through activation of other transcription factors. The role of WNT3-β-CATENIN signalling and mesodermal factor T in hPGC specification remains unknown but might be less significant.

Arrows and blunt-ended arrows depict positive and negative regulation respectively. Dotted lines indicate postulated regulations.

Figure 5: Epigenetic reprogramming in mouse and human PGCs.

a | Genome-wide DNA demethylation occurs in mPGCs during the onset of migration and early settlement at the genital ridge (~E7.5 to E13.5). Concurrently, mPGCs exhibit global chromatin

modification reorganization. In particular, H3K9me2, a repressive chromatin modification associated with silenced genes, becomes globally depleted. Since there is intimate crosstalk between H3K9me2 and DNA methylation pathways[179], loss of H3K9me2 in mPGCs may facilitate or be a consequence of global DNA demethylation. On the other hand, H3K9me3 levels remain high at pericentric heterochromatin-, while repressive H3K27me3 level become progressively enriched globally. H2A/H4R3me2s is transiently increased from E8.5 to E10.5, along with nuclear localization of the arginine methyltransferase PRMT5[131, 134]. The overall global epigenetic reprogramming dynamics in the human germline is similar to that in mPGCs, but with subtle differences. Notably, coupled loss of 5mC and H3K9me2 is also observed in hPGCs. However, H3K27me3 is only transiently enriched in hPGCs during migration and becomes low after arrival at the genital ridge[57, 180]. H2A/H4R3me2s levels stay constant throughout hPGC development (Tang WWC, unpublished data). Dotted lines indicate postulated dynamics in hPGCs.

b | Germline DNA demethylation erases parental epigenetic memories and drives germ cell differentiation. Allele-specific methylation at imprinting control regions (ICRs) are erased in PGCs and re-established in a sex-specific manner later in development. In XX PGCs, loss of DNA methylation and repressive H3K27me3 at the inactivated X-chromosome (labelled Xi) lead to X reactivation. This ensures each oocyte contains an active X-chromosome (labelled Xa) for early embryonic development after fertilization. In somatic cells, some germ cell-specific meiotic and genome defence genes (e.g. *DDX4*, *DAZL*, *SYCP3*, *PIWILs* and *KRAB-ZFPs*) are silenced by promoter DNA methylation. Demethylation of their CpG island (CGI) promoters in PGCs facilitates germ cell differentiation and maintains genome stability.

c | Despite global loss of DNA methylation, both evolutionary old (e.g. LINE-L2) and young (e.g. LTR-IAP in mice and LTR-HERVK in humans) retrotransposons remain repressed in PGCs. Repressive chromatin modifications are likely redistributed/retained at repetitive sequences to

safeguard genome stability. Notably, some evolutionarily young retrotransposons (e.g. LTR-IAP in mice and SVA in humans) and pericentromeric satellite repeats remain partially methylated which may contribute to their repression.

Glossary

Totipotency

The ability of a cell to give rise to all cell types (both embryonic and extraembryonic) of an organism.

Epiblast

Pluripotent cells derived from the inner cell mass of a blastocyst that give rise to all lineages of the embryo proper.

Allantois

A membranous sac that develops from the mesoderm (in mice) or hindgut endoderm (in humans) during early embryonic development. Allantois contributes to the formation of the umbilical cord and placenta.

Gastrulation

The developmental process in which the three germ layers (i.e. ectoderm, mesoderm and definitive endoderm) of the embryo are formed.

Primitive streak

A structure in the posterior end of the embryo where epiblast cells ingress to form the mesoderm and definitive endoderm. Formation of the primitive streak is the first visible sign of gastrulation.

Nodal Signalling

A signal transduction pathway that is essential for the formation of mesoderm and endoderm and axis determination in vertebrates. Nodal signalling is activated by transforming growth factor beta (TGF-β) family factors Activin and Nodal and transduced by SMAD2/3.

Trophoblast

The outermost layer of extraembryonic tissues that attaches the embryo to the uterine wall and forms the placenta.

Pluripotent states

Pluripotency refers to the ability of a cell to differentiate into any cell of the three germ layers in the embryo proper. The preimplantation epiblast represents a “naïve” pluripotent state, while the

postimplantation epiblast (poised for lineage differentiation) represents a “primed” pluripotent state.

Inner cell mass (ICM)

A compact mass of cells located at the embryonic pole of the blastocyst. ICM gives rise to the epiblast and the hypoblast, which forms the embryo proper and the yolk sac respectively.

Lineage specifiers

Transcription factors that direct competent cells to differentiate into a specific cell lineage.

Genomic imprinting

An epigenetic phenomenon that results in monoallelic gene expression in a parent-of-origin-dependent manner.

Bivalent promoter

A promoter simultaneously marked by both activating H3K4me3 and repressive H3K27me3 histone modifications. Genes with bivalent promoters are considered “poised” for activation when exposed to appropriate extrinsic signals.

Retrotransposons

DNA elements that can amplify themselves in the genome. During the process of retrotransposition, retrotransposon DNA is transcribed into RNA, then reverse transcribed into DNA, followed by insertion into a new genomic site.

KRAB-zinc finger proteins

The largest individual family of transcriptional repressors in mammals. KRAB-ZFPs contain DNA binding C2H2 zinc fingers and a Krüppel-associated box (KRAB) domain which interacts with KAP1 corepressor complex for epigenetic silencing.

Transgenerational epigenetic inheritance

Transmission of epigenetic information through the germline that affects phenotypic traits in more than one generation without changes in DNA sequence.

Highlighted references

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

A comprehensive signalling study which shows that BMP/pSMAD and WNT3 signallings are indispensable for mPGC specification from postimplantation epiblast during a restricted time-window.

Vincent, S.D. et al. The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. Development 132, 1315-25 (2005).

Ohinata, Y. et al. Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 436, 207-13 (2005).

The above papers show that *Prdm1* is the earliest lineage-restricted marker of mPGC and that it is indispensable for mPGC specification.

Yamaji, M. et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet 40, 1016-22 (2008).

This paper shows that *Prdm14* is critical for the upregulation of pluripotency genes and the initiation of epigenetic reprogramming during mPGC specification.

Weber, S. et al. Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biol Reprod 82, 214-23 (2010).

This paper shows that *Tfap2c*, together with *Prdm1*, repress mesodermal gene expression during mPGC specification.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell 146, 519-32 (2011).

This paper shows robust in vitro induction of mPGCLCs from ground-state mESCs via a postimplantation epiblast-like state.

Magnusdottir, E. et al. A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat Cell Biol 15, 905-15 (2013).

Nakaki, F. et al. Induction of mouse germ-cell fate by transcription factors in vitro. Nature 501, 222-6 (2013).

The above papers show that overexpression of *Prdm1*, *Tfap2c* and/or *Prdm14* is sufficient for induction of mPGC fate.

Nady, N. et al. ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. Elife 4 (2015).

Tu, S. et al. Co-repressor CBFA2T2 regulates pluripotency and germline development. Nature (2016).

The above papers show that co-repressor CBFA2T2 is a novel interactor of PRDM14 in regulation of pluripotency and mPGC specification.

Irie, N. et al. SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. Cell 160, 253-68 (2015).

This paper shows robust *in vitro* induction of hPGCLCs from germline competent hESCs and illustrates that SOX17 is essential for hPGC specification.

Guibert, S., Forne, T. & Weber, M. Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res 22, 633-41 (2012).

Seisenberger, S. et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 48, 849-62 (2012).

Hackett, J.A. et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science 339, 448-52 (2013).

The above papers reveal the detailed genome-wide DNA demethylation dynamics in mPGCs.

Walter, M., Teissandier, A., Perez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife 5 (2016).

This paper shows that repressive chromatin modifications repress retrotransposons and safeguard genome stability during chemical-induced global loss of DNA demethylation in mESCs.

Tang, W.W. et al. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. Cell 161, 1453-67 (2015).

Guo, F. et al. The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. Cell 161, 1437-52 (2015).

Through high quality RNA-Seq and BS-Seq analyses, the above papers reveal the transcriptional network and epigenetic reprogramming dynamics in hPGCs isolated from human embryos.

References

1. Ohinata, Y. et al. A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571-84 (2009).
2. Ohinata, Y. et al. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-13 (2005).
3. Yamaji, M. et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* **40**, 1016-22 (2008).
4. Weber, S. et al. Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biol Reprod* **82**, 214-23 (2010).
5. Magnusdottir, E. et al. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat Cell Biol* **15**, 905-15 (2013).
6. Magnusdottir, E. & Surani, M.A. How to make a primordial germ cell. *Development* **141**, 245-52 (2014).
7. Nakaki, F. et al. Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222-6 (2013).
8. Lawson, K.A. & Hage, W.J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* **182**, 68-84; discussion 84-91 (1994).
9. Seisenberger, S. et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* **48**, 849-62 (2012).
10. Hackett, J.A. et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448-52 (2013).
11. Guibert, S., Forne, T. & Weber, M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res* **22**, 633-41 (2012).
12. Hajkova, P. et al. Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23 (2002).
13. Lee, J. et al. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129**, 1807-17 (2002).
14. Sugimoto, M. & Abe, K. X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLoS Genet* **3**, e116 (2007).

15. Chuva de Sousa Lopes, S.M. et al. X chromosome activity in mouse XX primordial germ cells. *PLoS Genet* **4**, e30 (2008).

16. Tam, P.P., Zhou, S.X. & Tan, S.S. X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. *Development* **120**, 2925-32 (1994).

17. Monk, M. & McLaren, A. X-chromosome activity in foetal germ cells of the mouse. *J Embryol Exp Morphol* **63**, 75-84 (1981).

18. Seki, Y. et al. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134**, 2627-38 (2007).

19. Seki, Y. et al. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* **278**, 440-58 (2005).

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

21. De Felici, M. in Oogenesis (eds. Coticchio, G., Albertini, D.F. & De Santis, L.) 19-37 (Springer London, London, 2012).

22. Leitch, H.G., Tang, W.W. & Surani, M.A. Primordial germ-cell development and epigenetic reprogramming in mammals. *Curr Top Dev Biol* **104**, 149-87 (2013).

23. Fuss, A. Über extraregionare Geschlechtszellen bei einem menschlichen Embryo von 4 Wochen. *Anat Am* **39**, 407-9 (1911).

24. Witschi, E. Migration of the germ cells of human embryos from the yolk sac to the primitive gonadal folds. *Contrib Embryol* **32**, 67-80 (1948).

25. Extavour, C.G. & Akam, M. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* **130**, 5869-84 (2003).

26. Weismann, A., Parker, W.N. & Rönnfeldt, H. The germ-plasm : a theory of heredity (W. Scott, London, 1893).

27. Chatfield, J. et al. Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. *Development* **141**, 2429-40 (2014).

28. Ewen-Campen, B., Donoughe, S., Clarke, D.N. & Extavour, C.G. Germ cell specification requires zygotic mechanisms rather than germ plasm in a basally branching insect. *Curr Biol* **23**, 835-42 (2013).

29. Tam, P.P. & Loebel, D.A. Gene function in mouse embryogenesis: get set for gastrulation. *Nat Rev Genet* **8**, 368-81 (2007).

30. Aramaki, S. et al. A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev Cell* **27**, 516-29 (2013).

31. Ying, Y. & Zhao, G.Q. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev Biol* **232**, 484-92 (2001).

32. Lawson, K.A. et al. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13**, 424-36 (1999).

33. Ying, Y., Liu, X.M., Marble, A., Lawson, K.A. & Zhao, G.Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol Endocrinol* **14**, 1053-63 (2000).

34. de Sousa Lopes, S.M. et al. 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**, 1838-49 (2004).

35. Tremblay, K.D., Dunn, N.R. & Robertson, E.J. Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. *Development* **128**, 3609-21 (2001).

36. Chu, G.C., Dunn, N.R., Anderson, D.C., Oxburgh, L. & Robertson, E.J. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. *Development* **131**, 3501-12 (2004).

37. Chang, H. & Matzuk, M.M. Smad5 is required for mouse primordial germ cell development. *Mech Dev* **104**, 61-7 (2001).

38. Liu, P. et al. Requirement for Wnt3 in vertebrate axis formation. *Nat Genet* **22**, 361-5 (1999).

39. Brennan, J. et al. Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**, 965-9 (2001).

40. Beddington, R.S. & Robertson, E.J. Axis development and early asymmetry in mammals. *Cell* **96**, 195-209 (1999).

41. Behringer, R.R., Wakamiya, M., Tsang, T.E. & Tam, P.P. A flattened mouse embryo: leveling the playing field. *Genesis* **28**, 23-30 (2000).

42. Hopf, C., Viebahn, C. & Puschel, B. BMP signals and the transcriptional repressor BLIMP1 during germline segregation in the mammalian embryo. *Dev Genes Evol* **221**, 209-23 (2011).

43. Idkowiak, J., Weisheit, G., Plitzner, J. & Viebahn, C. Hypoblast controls mesoderm generation and axial patterning in the gastrulating rabbit embryo. *Dev Genes Evol* **214**, 591-605 (2004).

44. Valdez Magana, G., Rodriguez, A., Zhang, H., Webb, R. & Alberio, R. Paracrine effects of embryo-derived FGF4 and BMP4 during pig trophoblast elongation. *Dev Biol* **387**, 15-27 (2014).

45. Irie, N. et al. SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. *Cell* **160**, 253-68 (2015).

46. Kee, K., Gonsalves, J.M., Clark, A.T. & Pera, R.A. Bone morphogenetic proteins induce germ cell differentiation from human embryonic stem cells. *Stem Cells Dev* **15**, 831-7 (2006).

47. Sasaki, K. et al. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* **17**, 178-94 (2015).

48. Tam, P.P. & Zhou, S.X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* **178**, 124-32 (1996).

49. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-32 (2011).

50. Ying, Q.L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-23 (2008).

51. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol* **16**, 516-28 (2014).

52. Hayashi, K. et al. Offspring from oocytes derived from in vitro primordial germ cell-like cells in mice. *Science* **338**, 971-5 (2012).

53. Kee, K., Angeles, V.T., Flores, M., Nguyen, H.N. & Reijo Pera, R.A. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature* **462**, 222-5 (2009).

54. Wu, J. & Izpisua Belmonte, J.C. Dynamic Pluripotent Stem Cell States and Their Applications. *Cell Stem Cell* **17**, 509-25 (2015).

55. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J.H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol* **17**, 155-69 (2016).

56. Gafni, O. et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282-6 (2013).

57. Tang, W.W. et al. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* **161**, 1453-67 (2015).

58. Blakeley, P. et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3151-65 (2015).

59. Pastor, W.A. et al. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell* **18**, 323-9 (2016).

60. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254-69 (2014).

61. Sugawa, F. et al. Human primordial germ cell commitment in vitro associates with a unique PRDM14 expression profile. *EMBO J* (2015).

62. Kojima, Y. et al. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107-20 (2014).

63. Buecker, C. et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838-53 (2014).

64. Kurimoto, K. et al. Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. *Cell Stem Cell* **16**, 517-32 (2015).

65. Zylicz, J.J. et al. Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *Elife* **4** (2015).

66. Murakami, K. et al. NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature* **529**, 403-7 (2016).

67. Kurimoto, K. et al. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* **22**, 1617-35 (2008).

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

69. Vincent, S.D. et al. The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* **132**, 1315-25 (2005).

70. Grabole, N. et al. Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep* **14**, 629-37 (2013).

71. Martins, G. & Calame, K. Regulation and functions of Blimp-1 in T and B lymphocytes. *Annu Rev Immunol* **26**, 133-69 (2008).

72. Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. & Wysocka, J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat Struct Mol Biol* **18**, 120-7 (2011).

73. Yamaji, M. et al. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* **12**, 368-82 (2013).

74. Auman, H.J. et al. Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. *Development* **129**, 2733-47 (2002).

75. Werling, U. & Schorle, H. Transcription factor gene AP-2 gamma essential for early murine development. *Mol Cell Biol* **22**, 3149-56 (2002).

76. Nady, N. et al. ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife* **4**, e10150 (2015).

77. Tu, S. et al. Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature* (2016).

78. Fog, C.K., Galli, G.G. & Lund, A.H. PRDM proteins: important players in differentiation and disease. *Bioessays* **34**, 50-60 (2012).

79. Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J* **9**, 2185-95 (1990).

80. Yeom, Y.I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-94 (1996).

81. Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y. & Saitou, M. Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. *Biol Reprod* **75**, 705-16 (2006).

82. Sato, M. et al. Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech Dev* **113**, 91-4 (2002).

83. Leitch, H.G. et al. Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. *Development* **137**, 2279-87 (2010).

84. Matsui, Y., Zsebo, K. & Hogan, B.L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841-7 (1992).

85. Durcova-Hills, G., Tang, F., Doody, G., Tooze, R. & Surani, M.A. Reprogramming primordial germ cells into pluripotent stem cells. *PLoS One* **3**, e3531 (2008).

86. Kehler, J. et al. Oct4 is required for primordial germ cell survival. *EMBO Rep* **5**, 1078-83 (2004).

87. Yamaguchi, S. et al. Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells. *Development* **136**, 4011-20 (2009).

88. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-4 (2007).

89. Campolo, F. et al. Essential role of Sox2 for the establishment and maintenance of the germ cell line. *Stem Cells* **31**, 1408-21 (2013).

90. Yamaguchi, Y.L. et al. Sall4 is essential for mouse primordial germ cell specification by suppressing somatic cell program genes. *Stem Cells* **33**, 289-300 (2015).

91. Guo, F. et al. The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* **161**, 1437-52 (2015).

92. Perrett, R.M. et al. The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture. *Biol Reprod* **78**, 852-8 (2008).

93. de Jong, J. et al. Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications. *J Pathol* **215**, 21-30 (2008).

94. Hara, K. et al. Evidence for crucial role of hindgut expansion in directing proper migration of primordial germ cells in mouse early embryogenesis. *Dev Biol* **330**, 427-39 (2009).

95. Sarkar, A. & Hochedlinger, K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15-30 (2013).

96. Kanai-Azuma, M. et al. Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129**, 2367-79 (2002).

97. Kim, I., Saunders, T.L. & Morrison, S.J. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell* **130**, 470-83 (2007).

98. Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129-44 (2013).

99. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**, 101-6 (2008).

100. Niakan, K.K. et al. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev* **24**, 312-26 (2010).

101. Seguin, C.A., Draper, J.S., Nagy, A. & Rossant, J. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell* **3**, 182-95 (2008).

102. Stefanovic, S. et al. Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. *J Cell Biol* **186**, 665-73 (2009).

103. Aksoy, I. et al. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J* **32**, 938-53 (2013).

104. Niwa, H., Miyazaki, J. & Smith, A.G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-6 (2000).

105. Maruyama, M., Ichisaka, T., Nakagawa, M. & Yamanaka, S. Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *J Biol Chem* **280**, 24371-9 (2005).

106. Masui, S. et al. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**, 625-35 (2007).

107. Chia, N.Y. et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316-20 (2010).

108. Tsuneyoshi, N. et al. PRDM14 suppresses expression of differentiation marker genes in human embryonic stem cells. *Biochem Biophys Res Commun* **367**, 899-905 (2008).

109. Nady, N. et al. ETO family protein Mtgr1 mediates Prdm14 functions in stem cell maintenance and primordial germ cell formation. *Elife* **4** (2015).

110. Messerschmidt, D.M., Knowles, B.B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev* **28**, 812-28 (2014).

111. Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139**, 15-31 (2012).

112. Kobayashi, H. et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res* **23**, 616-27 (2013).

113. Sharif, J. et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908-12 (2007).

114. Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J* **32**, 340-53 (2013).

115. Ohno, R. et al. A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. *Development* **140**, 2892-903 (2013).

116. Arand, J. et al. Selective impairment of methylation maintenance is the major cause of DNA methylation reprogramming in the early embryo. *Epigenetics Chromatin* **8**, 1 (2015).

117. Hackett, J.A., Zylicz, J.J. & Surani, M.A. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends Genet* **28**, 164-74 (2012).

118. Yamaguchi, S. et al. Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res* **23**, 329-39 (2013).

119. Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460-4 (2013).

120. Dawlaty, M.M. et al. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev Cell* **24**, 310-23 (2013).

121. Yamaguchi, S. et al. Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443-7 (2012).

122. Gkountela, S. et al. DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell* **161**, 1425-36 (2015).

123. Gaskell, T.L., Esnal, A., Robinson, L.L., Anderson, R.A. & Saunders, P.T. Immunohistochemical profiling of germ cells within the human fetal testis: identification of three subpopulations. *Biol Reprod* **71**, 2012-21 (2004).

124. Kerr, C.L., Hill, C.M., Blumenthal, P.D. & Gearhart, J.D. Expression of pluripotent stem cell markers in the human fetal ovary. *Hum Reprod* **23**, 589-99 (2008).

125. Kerr, C.L., Hill, C.M., Blumenthal, P.D. & Gearhart, J.D. Expression of pluripotent stem cell markers in the human fetal testis. *Stem Cells* **26**, 412-21 (2008).

126. Wermann, H. et al. Global DNA methylation in fetal human germ cells and germ cell tumours: association with differentiation and cisplatin resistance. *J Pathol* **221**, 433-42 (2010).

127. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-20 (2013).

128. Tsumura, A. et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805-14 (2006).

129. Brinkman, A.B. et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**, 1128-38 (2012).

130. Walter, M., Teissandier, A., Perez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5** (2016).

131. Ancelin, K. et al. Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat Cell Biol* **8**, 623-30 (2006).

132. Ng, J.H. et al. In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev Cell* **24**, 324-33 (2013).

133. Liu, S. et al. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev* **28**, 2041-55 (2014).

134. Kim, S. et al. PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell* **56**, 564-79 (2014).

135. Guo, H. et al. The DNA methylation landscape of human early embryos. *Nature* **511**, 606-10 (2014).

136. Smith, Z.D. et al. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611-5 (2014).

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

138. Jeltsch, A. & Jurkowska, R.Z. New concepts in DNA methylation. *Trends Biochem Sci* **39**, 310-8 (2014).

139. Wolf, G., Greenberg, D. & Macfarlan, T.S. Spotting the enemy within: Targeted silencing of foreign DNA in mammalian genomes by the Kruppel-associated box zinc finger protein family. *Mob DNA* **6**, 17 (2015).

140. Jacobs, F.M. et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**, 242-5 (2014).

141. Li, X. et al. A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev Cell* **15**, 547-57 (2008).

142. Mackay, D.J. et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet* **40**, 949-51 (2008).

143. Williams, Z. et al. Discovery and Characterization of piRNAs in the Human Fetal Ovary. *Cell Rep* **13**, 854-63 (2015).

144. Baudat, F. et al. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836-40 (2010).

145. Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* **13**, 153-62 (2012).

146. Bohacek, J. & Mansuy, I.M. Molecular insights into transgenerational non-genetic inheritance of acquired behaviours. *Nat Rev Genet* **16**, 641-52 (2015).

147. Heard, E. & Martienssen, R.A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95-109 (2014).

148. Surani, M.A., Barton, S.C. & Norris, M.L. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548-50 (1984).

149. McGrath, J. & Solter, D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179-83 (1984).

150. Radford, E.J. et al. In utero effects. In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345**, 1255903 (2014).

151. Shea, J.M. et al. Genetic and Epigenetic Variation, but Not Diet, Shape the Sperm Methylome. *Dev Cell* **35**, 750-8 (2015).

152. Chen, Q. et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397-400 (2016).

153. Sharma, U. et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391-6 (2016).

154. Johnson, A.D. & Alberio, R. Primordial germ cells: the first cell lineage or the last cells standing? *Development* **142**, 2730-9 (2015).

155. Zhou, Q. et al. Complete Meiosis from Embryonic Stem Cell-Derived Germ Cells In Vitro. *Cell Stem Cell* **18**, 1-11 (2016).

156. Zhang, X. et al. Pax6 is a human neuroectoderm cell fate determinant. *Cell Stem Cell* **7**, 90-100 (2010).

157. Rossant, J. & Tam, P.P. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701-13 (2009).

158. De Paepe, C., Krivega, M., Cauffman, G., Geens, M. & Van de Velde, H. Totipotency and lineage segregation in the human embryo. *Mol Hum Reprod* **20**, 599-618 (2014).

159. Niakan, K.K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev Biol* **375**, 54-64 (2013).

160. O'Leary, T. et al. Tracking the progression of the human inner cell mass during embryonic stem cell derivation. *Nat Biotechnol* **30**, 278-82 (2012).

161. Bedzhov, I. & Zernicka-Goetz, M. Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* **156**, 1032-44 (2014).

162. Moore, K.L., Persaud, T.V.N. & Torchia, M.G. The developing human : clinically oriented embryology (Elsevier/Saunders, Philadelphia, PA, 2013).

163. Hackett, J.A. & Surani, M.A. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell* **15**, 416-30 (2014).

164. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6 (1981).

165. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8 (1981).

166. Brons, I.G. et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-5 (2007).

167. Tesar, P.J. et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-9 (2007).

168. Huang, Y., Osorno, R., Tsakiridis, A. & Wilson, V. In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Rep* **2**, 1571-8 (2012).

169. Toyooka, Y., Tsunekawa, N., Akasu, R. & Noce, T. Embryonic stem cells can form germ cells in vitro. *Proc Natl Acad Sci U S A* **100**, 11457-62 (2003).

170. Hubner, K. et al. Derivation of oocytes from mouse embryonic stem cells. *Science* **300**, 1251-6 (2003).

171. Hayashi, K. & Surani, M.A. Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. *Development* **136**, 3549-56 (2009).

172. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-7 (1998).

173. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487-92 (2009).

174. Vallier, L., Alexander, M. & Pedersen, R.A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118**, 4495-509 (2005).

175. Son, M.Y., Choi, H., Han, Y.M. & Cho, Y.S. Unveiling the critical role of REX1 in the regulation of human stem cell pluripotency. *Stem Cells* **31**, 2374-87 (2013).

176. Clark, A.T. et al. Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. *Hum Mol Genet* **13**, 727-39 (2004).

177. Rao, J. et al. Stepwise Clearance of Repressive Roadblocks Drives Cardiac Induction in Human ESCs. *Cell Stem Cell* **18**, 341-53 (2016).

178. Lin, I.Y. et al. Suppression of the SOX2 neural effector gene by PRDM1 promotes human germ cell fate in embryonic stem cells. *Stem Cell Reports* **2**, 189-204 (2014).

179. Du, J., Johnson, L.M., Jacobsen, S.E. & Patel, D.J. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* **16**, 519-32 (2015).

180. Gkountela, S. et al. The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat Cell Biol* **15**, 113-22 (2013).