nature cell biology

# The ontogeny of cKIT$^+$ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and *in vitro* differentiation

**Sofia Gkountela[1,2], Ziwei Li[1,2], John J. Vincent[1,2,3], Kelvin X. Zhang[4], Angela Chen[5], Matteo Pellegrini[1] and Amander T. Clark[1,2,3,6,7]**

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

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

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

Human germ line development begins with the formation of PGCs that express the tyrosine kinase receptor cKIT (refs 4–10). Very little is known about the developmental progression (ontogeny) of cKIT$^+$ PGCs; however, on the basis of the mouse model, it is clear that PGCs must undergo whole-genome epigenetic reprogramming to remove cytosine methylation from imprinted genes and restore totipotency[11–14]. Given the fundamental role of epigenetic reprogramming in the germ line, it is essential to characterize reprogramming in human PGCs because mouse and human genomes are separated by ~170 million years and diverse strategies may have evolved to execute it. Once the major molecular landmarks of human PGC reprogramming are known, we propose that this information will be critical to assessing PGC differentiation and reprogramming *in vitro*, or identifying bottlenecks that must be overcome to generate a functional germ line from hPSCs.

By evaluating 134 human embryonic and fetal gonadal samples from 6 to 20 developmental weeks, we provide the first comprehensive tran-

[1]Department of Molecular Cell and Developmental Biology, University of California Los Angeles, Los Angeles, California 90095, USA. [2]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, California 90095, USA. [3]Molecular Biology Institute, University of California Los Angeles, Los Angeles, California 90095, USA. [4]Department of Biological Chemistry, Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, California 90095, USA. [5]Obstetrics & Gynecology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. [6]Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, California 90095, USA.
[7]Correspondence should be addressed to A.T.C. (e-mail: clarka@ucla.edu)

Received 9 July 2012; accepted 1 November 2012; published online 16 December 2012; DOI: 10.1038/ncb2638

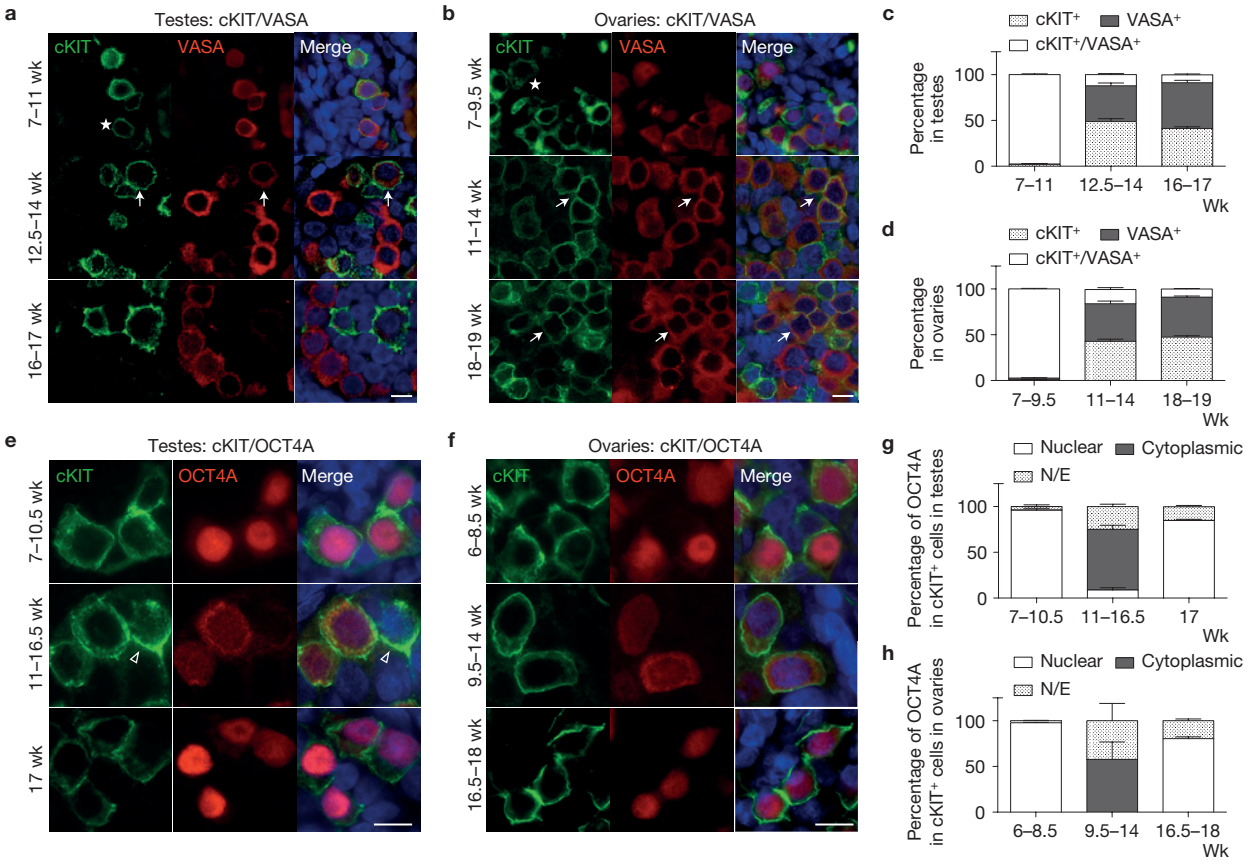


**Figure 1** The dynamics of cKIT, OCT4A and VASA expression in the fetal gonad. (**a**,**b**) Representative immunofluorescence micrographs of cKIT with VASA at the developmental weeks indicated. Stars indicate cKIT dim cells. (**a**) A representative example of a 10-week testis sample drawn from the group designated 7–11 wk (5 samples analysed in total), a 13.5-week testis for 12.5–14 wk (3 samples) and a 16-week testis for 16–17 wk (2 samples). (**b**) A 7-week ovary for 7–9.5 wk (4 samples), an 11-week ovary for 11–14 wk (3 samples) and an 18-week ovary for 18–19 wk (2 samples). (**c**,**d**) Quantifcation of cKIT$^+$, VASA$^+$ and cKIT$^+$/VASA$^+$cells (arrows in **a**,**b**). (**c**) In testes, $n = 9$ optic fields were counted at 7–11 wk (collected from 5 samples), $n = 9$ optic fields at 12.5–14 wk (3 samples) and $n = 7$ optic fields at 16–17 wk (2 samples). (**d**) In ovaries, $n = 7$ optic fields were counted at 7–9.5 wk (4 samples), $n = 8$ optic fields at 11–14 wk (3 samples) and $n = 7$ optic fields at 18–19 wk (3 samples). (**e**,**f**) Representative immunofluorescence micrographs of cKIT with OCT4A at the developmental weeks indicated. (**e**) A 10-week testis for 7–10.5 wk (5 samples), a 13.5-week testis for 11–16.5 wk (4 samples) and a 17-week testis (1 sample). Arrowhead indicates a cKIT$^+$ OCT4A-negative cell. (**f**) An 8-week ovary for 6–8.5 wk (3 samples), an 11-week ovary for 9.5–14 wk (3 samples) and an 18-week ovary for 16.5–18 wk (2 samples). (**g**,**h**) Quantification of nuclear or cytoplasmic localization of OCT4A in cKIT$^+$ cells. (**g**) In testes, $n = 6$ optic fields were counted at 7–10.5 wk (5 samples), $n = 8$ optic fields at 11–16.5 wk (4 samples) and $n = 6$ optic fields at 17 wk (1 sample). (**h**) In ovaries, $n = 6$ optic fields were counted at 7–8.5 wk (3 samples), $n = 8$ optic fields at 9.5–14 wk (3 samples) and $n = 9$ optic fields at 16.5–18 wk (2 samples). For immunofluorescence microscopy, nuclei were counterstained with DAPI (blue); scale bars, 10 μm. For quantification, on average 40 cells per optic field were counted using a 40× objective. All data are expressed as mean ± s.e.m. of $n = total$ number of optic fields per age group as indicated. Wk (wk), week; N/E, not expressed.

scriptional and epigenetic roadmap of human cKIT$^+$ PGCs in testes and ovaries and pinpoint the timing of major epigenetic events including whole-genome reprogramming and initiation of imprint erasure. Using the endogenous human cKIT$^+$ PGCs as a reference, we can now more accurately interpret the identity of the cKIT$^+$ subpopulation of PGCs acquired with *in vitro* differentiation from hESCs. Our results clearly demonstrate that single-cell analysis at both RNA and protein level is critical to defining PGC identity *in vitro*, and indisputably shows that established hESC lines are not equivalent to human PGCs.

## RESULTS

### cKIT$^+$ PGCs undergo molecular progression with fetal development

Temporal and spatial expression of cKIT in fetal testes and ovaries from 7 to 19 weeks of development was evaluated by immunofluorescence microscopy together with the evolutionarily conserved germ cell marker VASA. All testes samples procured had characteristic seminiferous cords by histology, indicating that sex determination had been initiated[15] (Supplementary Fig. S1). We identified cKIT on the surface of all VASA$^+$ cells in testes from 7 to 11 weeks and ovaries from 7 to 9.5 weeks (Fig. 1a,b and Supplementary Fig. S2a). However, from 12.5 weeks in testes, and 11 weeks in ovaries, cKIT and VASA protein expression becomes uncoupled, with only 10% of cKIT$^+$ cells co-expressing VASA (arrows in Fig. 1a,b, quantified in Fig. 1c,d and Supplementary Fig. S2b). On uncoupling, the ratio of single cKIT$^+$ to single VASA$^+$ cells was 1:1. We also evaluated SSEA1, and found that although PGCs are SSEA1$^+$ in fetal testes at the 'common PGC progenitor stage' and after cKIT/VASA uncoupling, SSEA1 alone is not specific for the human germ line because it was also expressed on cKIT- and VASA-negative cells (not germ cells; Supplementary Fig. S3a,b).

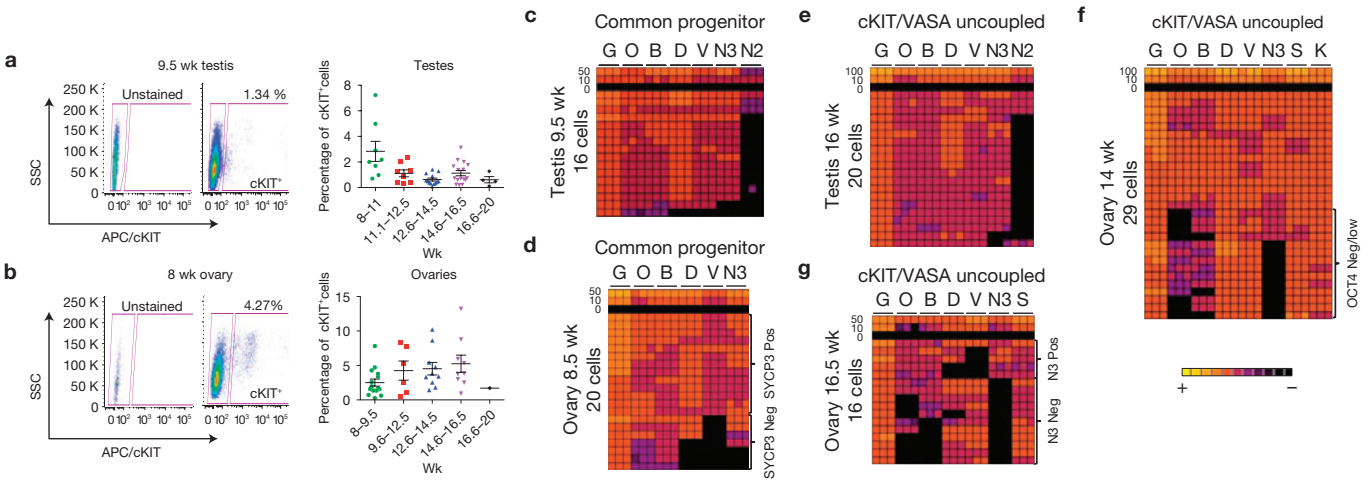


**Figure 2** Molecular characterization of cKIT$^+$ PGCs from 7 to 20 developmental weeks. (**a**,**b**) Gating strategy for sorting cKIT$^+$ cells with an APC-conjugated anti-human cKIT primary antibody against side scatter (SSC) for a 9.5-week testis (**a**) and an 8-week ovary (**b**). Also shown is the percentage of cKIT$^+$ cells sorted from the live fraction of testes in **a** and ovaries in **b** at 8–20 developmental weeks (wk). Each data point represents a single sample (biological replicate). All data are represented as mean ± s.e.m. (**c**–**g**) Heat map of *GAPDH* (G), *OCT4* (O) , *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3), *cKIT* (K) *NANOS2* (N2) and *SYCP3* (S) in triplicate (columns) in 100, 50, 10, 0 or single sorted cKIT$^+$ cells (rows).

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

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

To determine the molecular identity of cKIT$^+$ PGCs, we performed single-cell analysis with five PGC signature genes including *OCT4*, *BLIMP1*, *DAZL*, *VASA* and *NANOS3* using FACS, followed by microfluidic qRT–PCR at the common progenitor and cKIT/VASA uncoupled stage (Fig. 2c–g). We also confirmed expression of cKIT in individual cells (Fig. 2f). Our single-cell approach was first validated in HEK293T cells (Supplementary Fig. S4e–h). At the common PGC progenitor stage, 14/16 cKIT$^+$ cells in the testis and 13/20 cKIT$^+$ cells in the ovary coordinately expressed the five PGC signature genes (Fig. 2c,d). However, in the ovary 7/20 cKIT$^+$ cells did not express *VASA* and/or *DAZL* at this stage and instead were *OCT4/BLIMP1* double positive (O/B) or *OCT4/BLIMP1/NANOS3* triple positive (O/B/N3). In testes, *NANOS2* expression was also evaluated and found in <20% of cKIT$^+$ cells in the common progenitor and this was maintained on cKIT/VASA uncoupling in the cKIT$^+$ cell (Fig. 2c,e). In the ovary during the uncoupled stage when OCT4A is either in the cytoplasm or no longer expressed, *NANOS3* messenger RNA is also no longer expressed in a fraction of cells, and these *NANOS3*-negative cells correlated with no or low levels of *OCT4* and *BLIMP1* (Fig. 2f). At 16.5 weeks, when OCT4A is again localized to the nucleus or not expressed, ovarian cKIT$^+$ cells become even more

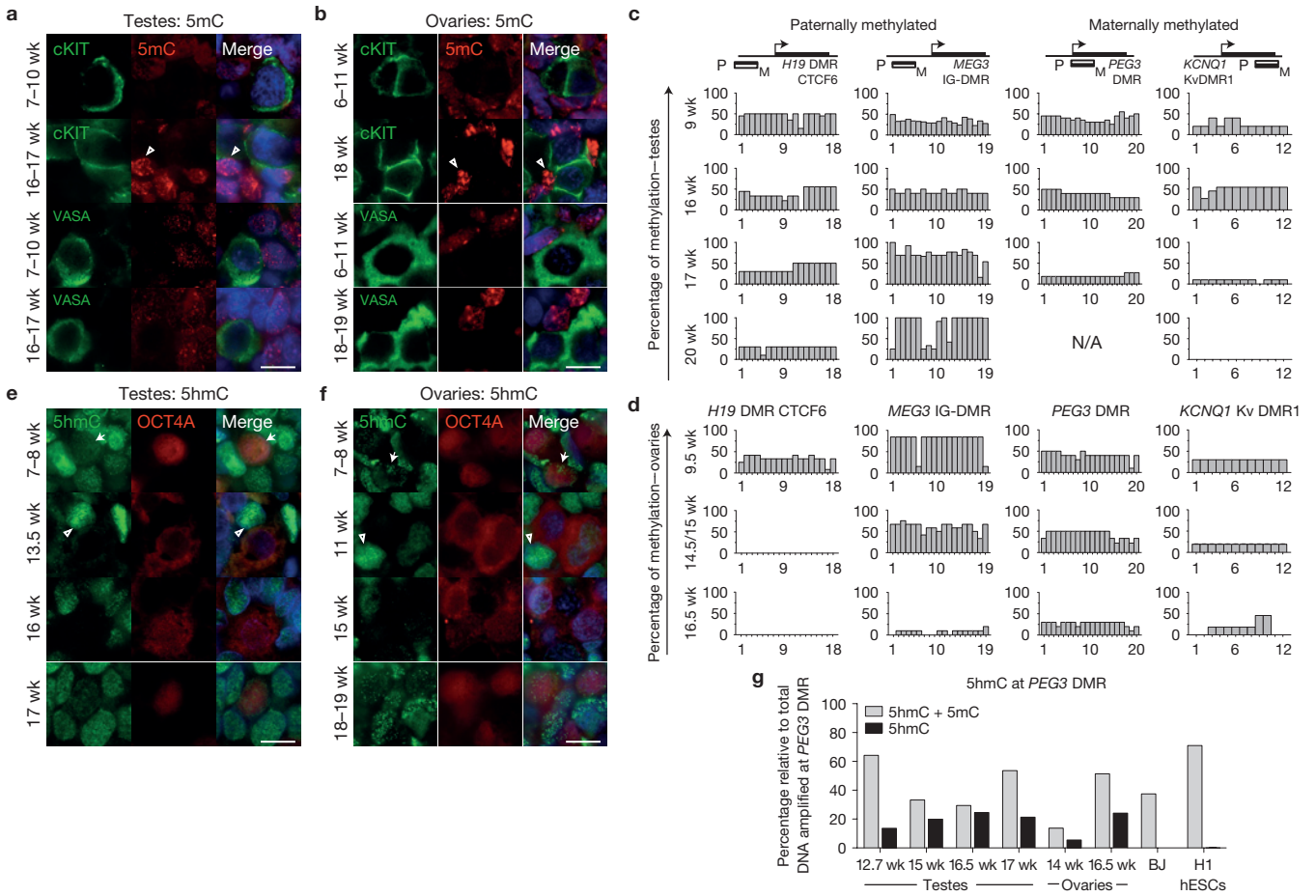


**Figure 3** Global loss of 5mC precedes loss of 5hmC. (**a**,**b**) Representative immunofluorescence micrographs of 5mC with cKIT or with VASA in testes (**a**) and ovaries (**b**) at the developmental weeks indicated. Open arrowheads indicate 5mC signal in somatic cells. (**a**) A 10-week testis for 7–10 wk ($n = 3$) and a 17-week testis for 16–17 wk ($n = 3$). (**b**) An 8-week ovary for 6–11 wk ($n = 3$) and an 18-week ovary for 18–19 wk ($n = 2$). (**c**,**d**) BS-PCR analysis of *H19*, *MEG3*, *PEG3* and *KCNQ1* in cKIT$^+$PGCs sorted from testes at 9 weeks ($n = 2$) and at 16, 17 and 20 weeks (**c**), and ovaries at 9.5, 14.5, 15 and 16.5 weeks (**d**). (**e**,**f**) Representative immunofluorescence micrographs of 5hmC with OCT4A in testes (**e**) and ovaries (**f**) at the developmental stages indicated in weeks. Arrows indicate 5hmC signal in PGCs; open arrowheads indicate 5hmC signal in somatic cells. (**e**) An 8-week testis for 7–8 wk ($n = 2$), a 13.5-week testis ($n = 1$), a 16-week testis ($n = 2$) and a 17-week testis ($n = 1$). (**f**) An 8-week ovary for 7–8 wk ($n = 2$), an 11-week ovary, ($n = 1$), a 15-week ovary ($n = 1$) and an 18-week ovary for 18–19 wk ($n = 2$). (**g**) CGRA of the *PEG3* DMR showing the percentage of total methylation (5mC + 5hmC) or 5hmC alone, relative to total amplified DNA (uncut) at the *PEG3* DMR. DNA from BJ fibroblast and H1 hESCs was used as a negative control (for each, $n = 2$ biological replicates). For immunofluorescence microscopy analysis, nuclei were counterstained with DAPI (blue). Scale bars, 10 μm. All data are represented as mean ± s.e.m. N/A, not amplified; wk, week.

heterogeneous, most notably involving loss of *NANOS3*, *OCT4* and *BLIMP1* mRNA in some cells, with *DAZL* and *VASA* being absent in others (Fig. 2g). *SYCP3* was used to indicate meiotic potential, and was expressed in every cell at 14 and 16.5 weeks (Fig. 2f,g). Furthermore, at 16.5 weeks, *SYCP3* and *VASA* mRNA expression levels were significantly enriched in the *NANOS3*-negative population (Supplementary Fig. S4d). Despite *SYCP3* mRNA expression in every cell at a single-cell level, on the protein level, only VASA$^+$ cells are immunopositive for SYCP3 in the fetal ovary from 14 weeks and not cKIT$^+$ (Supplementary Fig. S3e), indicating that VASA$^+$ cells are the first to acquire meiotic potential.

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

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

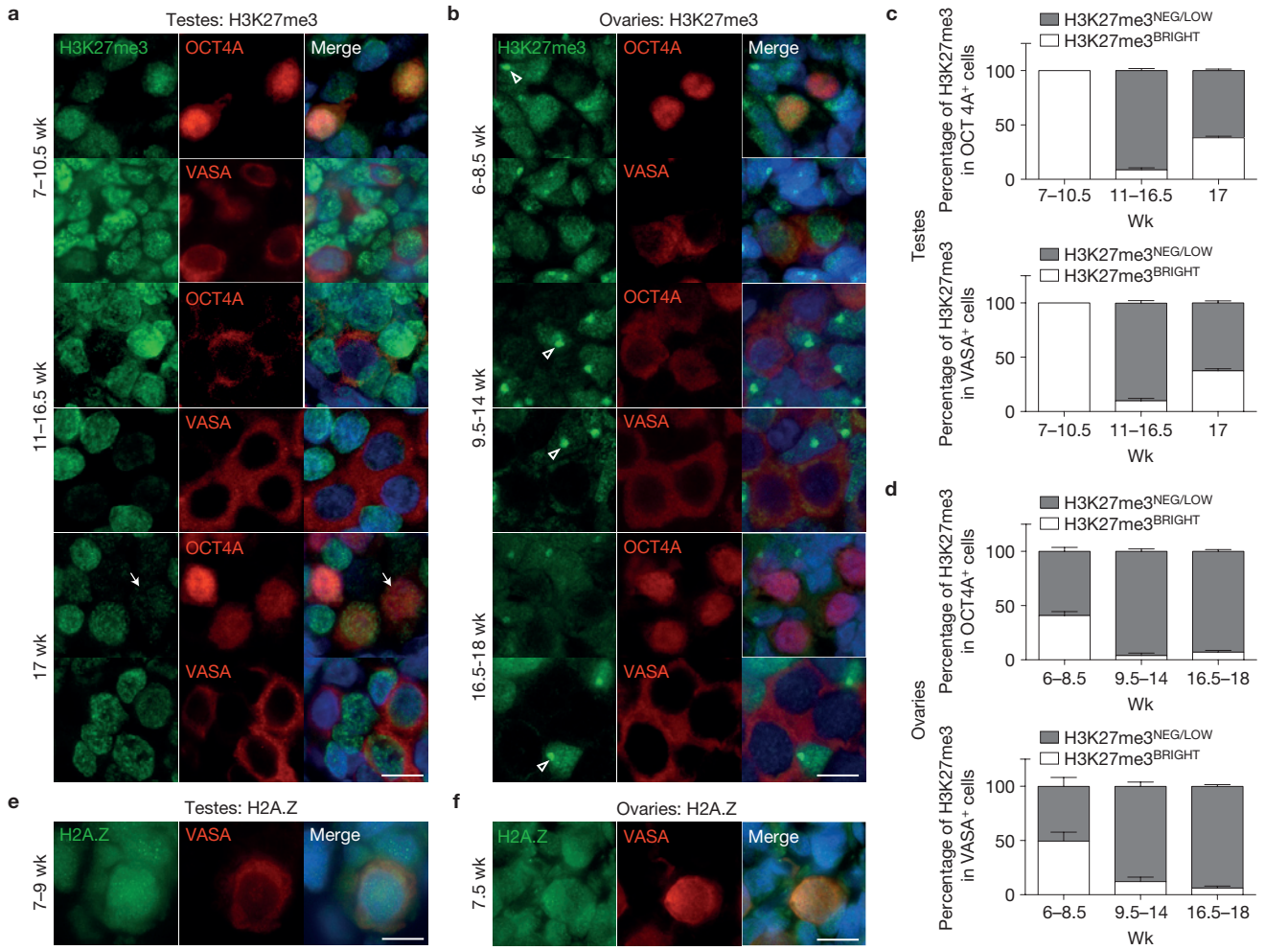


**Figure 4** Epigenetic reprogramming of H3K27me3 and H2A.Z occurs in the common PGC progenitor. (**a**,**b**) Representative immunofluorescence micrographs of H3K27me3 with OCT4A or VASA in testes from 7 to 17 weeks (**a**), and ovaries from 6 to 18 weeks (**b**). (**a**) A representative example of a 10.5-week testis sample drawn from the group designated r 7–10.5 wk (3 samples analysed in total), a 16.5-week testis for 11–16 wk (7 samples) and a 17-week testis (1 sample). Arrow indicates PGC nucleus with H3K27me3Low levels at 17 weeks relative to the intensity of staining in the somatic neighbours in the same section. (**b**) A 7-week ovary for 6–8.5 wk (3 samples), an 11-week ovary for 9.5–14 wk (4 samples) and an 18-week ovary for 16.5–18 wk (3 samples). Open arrowhead indicates strong H3K27me3 accumulation that is indicative of X chromosome inactivation[42]. (**c**,**d**) Quantification of H3K27me3 in OCT4A$^+$ or VASA$^+$ germ cells in testes (**c**) and ovaries (**d**), at the developmental ages indicated. (**c**) In testes for quantification in OCT4A$^+$, $n=6$ optic fields were counted at 7–10.5 wk (collected from 3 samples), $n=14$ optic fields at 11–16.5 wk (7 samples) and $n=6$ optic fields at 17 wk (1 sample). (**d**) In ovaries, for quantification in OCT4A$^+$, $n=6$ optic fields were counted at 6–8.5 wk (3 samples), $n=10$ optic fields at 9.5–14 wk (4 samples) and $n=8$ optic fields at 16.5–18 wk (3 samples). For quantification in VASA$^+$, $n=4$ optic fields were counted at 6–8.5 wk (3 samples), $n=10$ optic fields at 9.5–14 wk (4 samples) and $n=7$ optic fields at 16.5–18 wk (3 samples). (**e**,**f**) Representative immunofluorescence micrographs of H2A.Z with VASA in testes from 7 to 9 weeks (**e**), and ovary at 7.5 weeks (**f**). (**e**) A 9-week testis for 7–9 wk ($n=2$). (**f**) A 7.5-week ovary ($n=2$). For immunofluorescence microscopy, nuclei were counterstained with DAPI (blue). Scale bars, 10 μm. For quantification, on average 40 cells per optic field were counted using a 40× objective. All data are expressed as mean ± s.e.m. of $n=$ total number of optic fields per age group as indicated. Wk (wk), week.

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

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

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

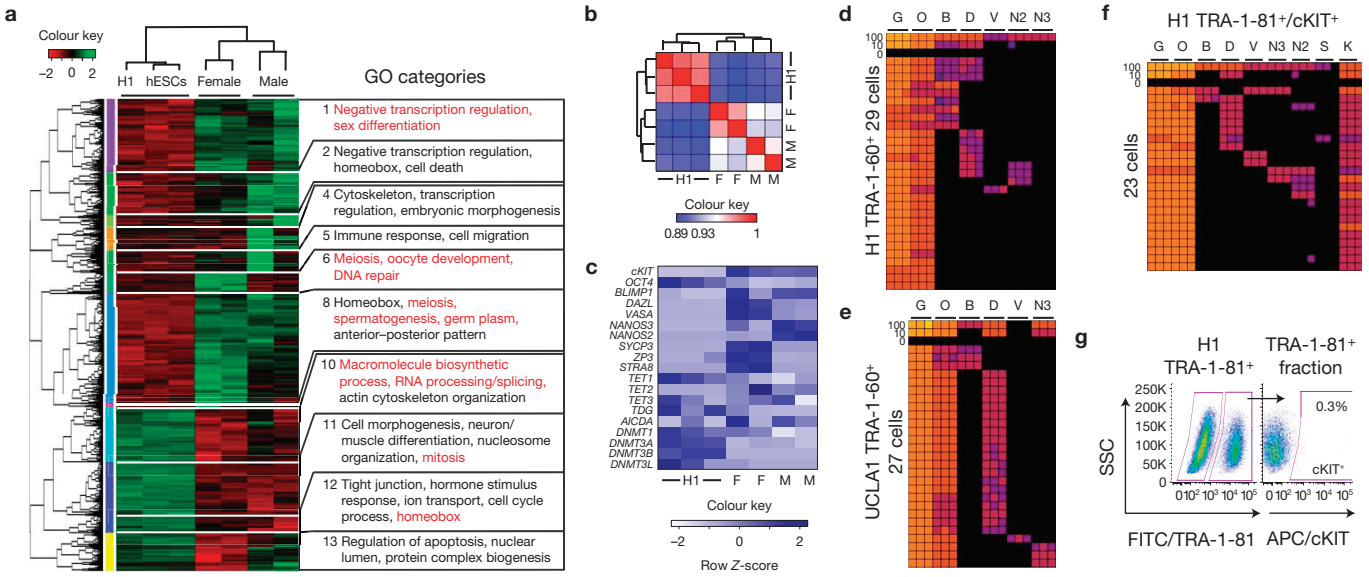


**Figure 5** RNA-Seq reveals the transcriptional identity of cKIT$^+$ PGCs. Single-cell analysis of hESCs shows stochastic expression of germ line genes. (**a**) Heat map of 5,455 differentially expressed genes ($P < 0.05$) in at least one of three comparisons (male cKIT$^+$ versus H1 hESCs; female cKIT$^+$ versus H1 hESCs; male cKIT$^+$ versus female cKIT$^+$). The enriched GO terms in the 13 resulting clusters are shown. (**b**) Heat map of Pearson correlation coefficient scores between hESCs and cKIT$^+$ male and female PGCs. (**c**) Heat map of FPKM values for selected genes in hESCs and cKIT$^+$ male and female PGCs. M, male; F, female. (**d**,**e**) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3) and *NANOS2* (N2) for H1 hESCs in triplicate (columns) in 100, 10, 0 or single TRA-1-60$^+$ cells (rows) sorted from H1 (**d**) and UCLA1 (**e**) hESCs. (**f**) Heat map as in **d** and **e** plus *cKIT* (K) for TRA-1-81$^+$/cKIT$^+$ H1 hESCs. (**g**) Gating strategy to sort TRA-1-81$^+$/cKIT$^+$ cells from the H1 hESC line. cKIT$^+$ cells are gated from the TRA-1-81$^+$ fraction, using a FITC secondary antibody against side scatter (SSC).

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

In the mouse, gonadal epigenetic reprogramming of PGCs and imprint erasure occurs from embryonic day (E)11.5 to E12.5 coincident with global changes in chromatin, including a transient loss of tri-methylation of Lys 27 on histone H3 (H3K27me3) and a permanent loss of the histone variant H2A.Z (ref. 12). In humans, using immunofluorescence microscopy we show that H3K27me3 is enriched in the nucleus of common PGC progenitors in testes from 7 to 10.5 weeks (Fig. 4a,c). However, at 11 weeks, the endpoint of the common progenitor stage, H3K27me3 is at or below the level of detection in most OCT4A$^+$ and VASA$^+$ PGCs (Fig. 4a,c). Interestingly, at 17 weeks in testes, H3K27me3 is again observed in the nucleus of ~38% OCT4A$^+$ and VASA$^+$ PGCs. In ovaries, H3K27me3 is absent in 50–60% of common PGC progenitors at 6–8.5 weeks (Fig. 4b,d), after which all PGCs are negative for H3K27me3 (Fig. 4b,d). Similarly, H2A.Z is enriched in the nucleus of common-progenitor-stage PGCs at 7–9 weeks in the testis and 7.5 weeks in the ovary (Fig. 4e,f). However, at the end of the common progenitor stage all PGCs become devoid of H2A.Z until around 17 weeks when H2A.Z reappears in the nucleus of a few VASA$^+$ cells in both sexes (Supplementary Fig. S5b,c).

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

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

Given that 5hmC was detected at 16–16.5 weeks by either CGRA and/or immunofluorescence microscopy in cKIT$^+$ PGCs, we also examined the expression of the ten-eleven translocation (*TET*) genes, which are responsible for converting 5mC to 5hmC (refs 19–21; Fig. 5c). All three *TET* family members *(TET1–3)* are expressed by male and female PGCs, with a significant enrichment of *TET2*, and reduced expression of *TET1* relative to H1 hESCs. We also evaluated the DNA methyltransferases (DNMTs) *DNMT1*, *DNMT3A*, *DNMT3B* and *DNMT3L*. Male but not female PGCs exhibited a significant decrease in *DNMT1* relative to H1 hESCs. Furthermore, all PGC samples had reduced levels of expression of *DNMT3A* and *DNMT3B* relative to H1. *AICDA* (also known as *AID*) and *TDG* were expressed at variable levels in H1 and also in PGCs of both sexes.

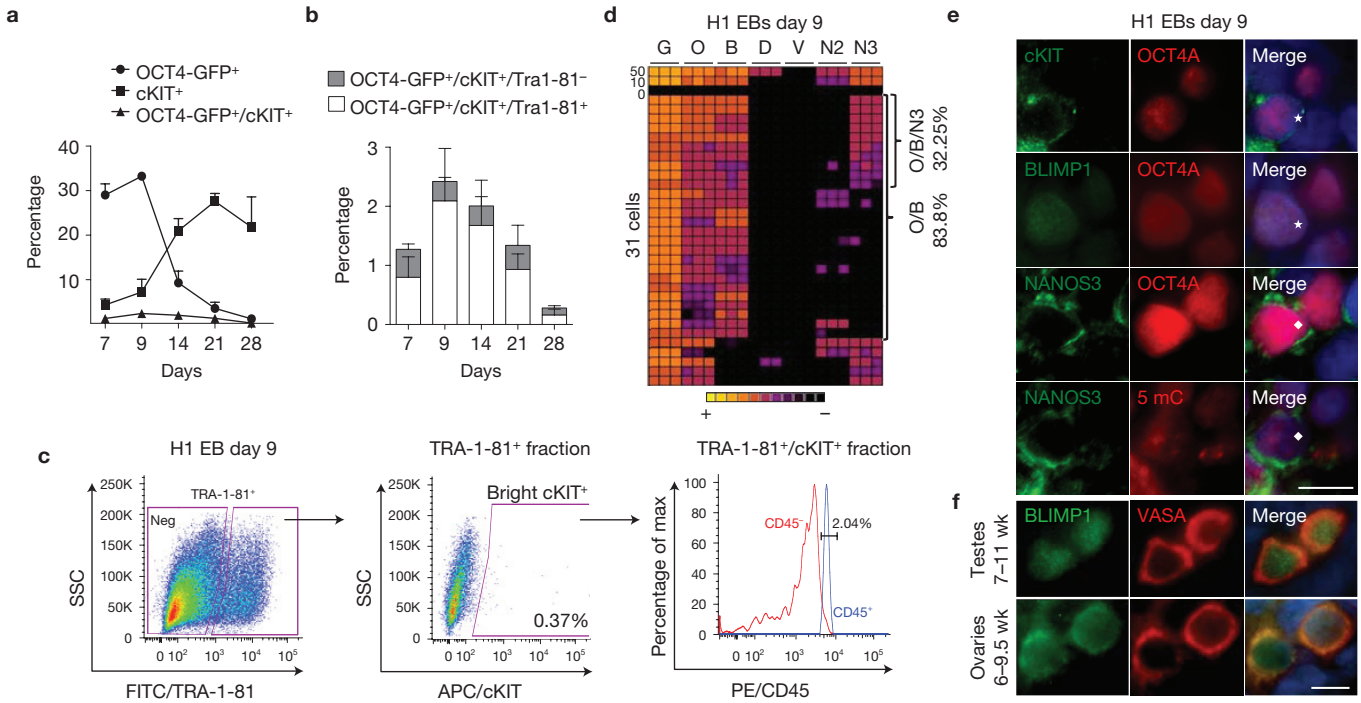


**Figure 6** *In vitro* hESC differentiation generates rare germ line progenitors that are cKIT/TRA-1-81 positive. (**a**) Percentage of OCT4–GFP$^+$, cKIT$^+$ and total OCT4–GFP$^+$/cKIT$^+$ generated on adherent differentiation for the indicated time points ($n=3$). (**b**) Percentage of TRA-1-81$^+$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

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

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

Next, PGC differentiation was evaluated for up to 16 days by serum-induced differentiation of hESCs: as embryoid bodies with, and without, BMP4 addition; adherent monolayer differentiation on growth-factor-reduced matrigel (GFR M/G); differentiation on human fetal gonadal stromal cells (hFGSC); and a combination of embryoid body differentiation, followed by plating embryoid bodies on hFGSCs (Figs 6 and 7). First we verified that TRA-1-81 faithfully reports OCT4 expression on hESC differentiation for 28 days by flow cytometry using the H1 OCT4–GFP line created by homologous recombination[27] (Fig. 6a,b). Using the gating strategy shown in Fig. 6c (H1 day-9 embryoid bodies), we sorted the brightest TRA-1-81$^+$/cKIT$^+$ cells on *in vitro* differentiation averaging 0.3% of the live population, with no significant difference in the percentage of positive cells when comparing differentiation strategy or length of time in differentiation. Our data show that >97% of TRA-1-81$^+$/cKIT$^+$ differentiated cells are negative for CD45, excluding the possibility of contamination with cKIT$^+$ haematopoietic progenitors[28] (Fig. 6c). Single-cell analysis of cKIT$^+$/TRA-1-81$^+$ cells sorted from embryoid bodies at day 9 revealed a significant increase in the proportion of cells expressing *BLIMP1* (Fig. 6d). We show that the identity of putative PGCs was heterogeneous being either O/B double positive, or O/B/N3 triple positive with no co-expression of *DAZL* or *VASA* (Figs 6d and 7c). Although rare O/B and O/B/N3 single cells were identified in the undifferentiated state (Fig. 5d–f), differentiation resulted in a clear enrichment for both O/B and O/B/N3 cell types. In particular, embryoid body differentiation for 9 days yielded 9-fold enrichment in O/B and 7.5-fold enrichment in O/B/N3 cells relative to cKIT$^+$/TRA-1-81$^+$ self-renewing hESCs (quantified in Fig. 7a,b). Immunofluorescence microscopy of day-9 embryoid bodies verified that cKIT$^+$ cells co-expressed OCT4A, exhibited nuclear localization of BLIMP1 and expressed NANOS3 in the cytoplasm, sim-

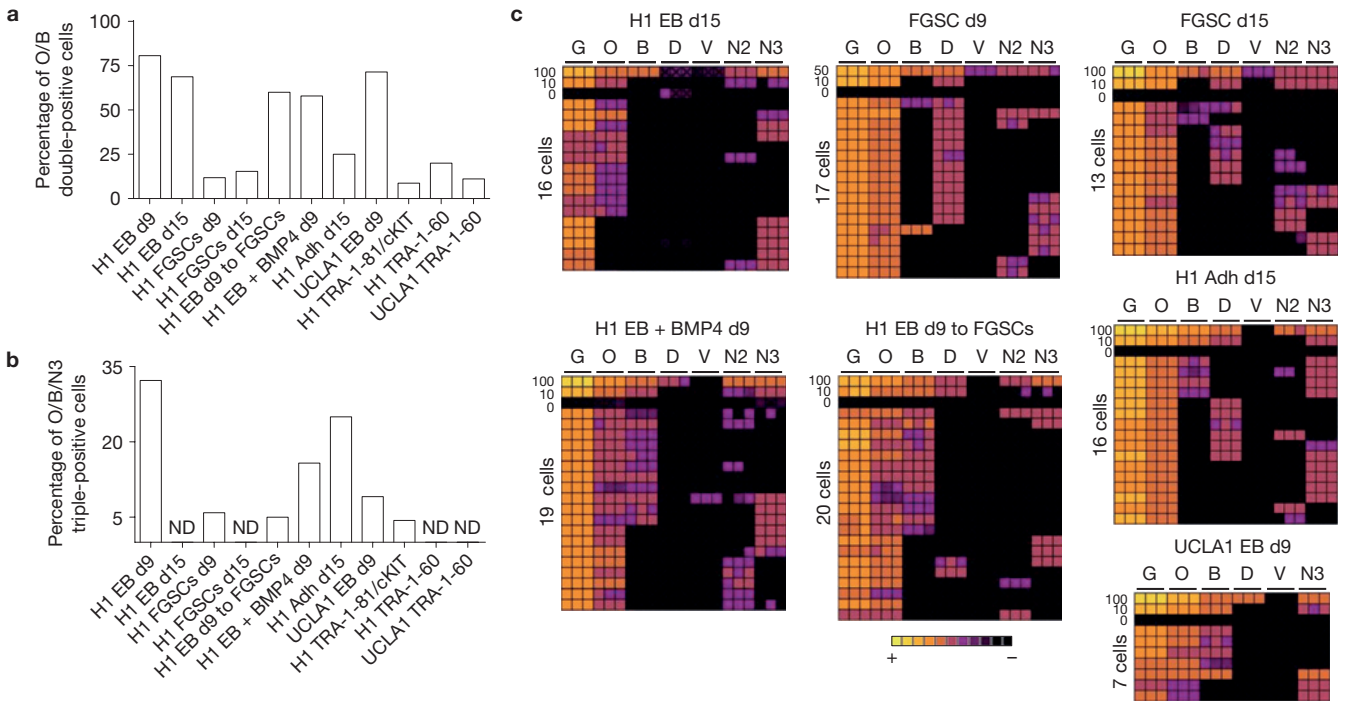


**Figure 7** *In vitro* PGC differentiation from hESCs using five alternative differentiation techniques. (**a**,**b**) Comparison of the percentage of O/B double-positive single cells (**a**) and O/B/N3 triple-positive cells (**b**) within the TRA-1-81$^+$/cKIT$^+$ sorted fraction from five alternative differentiation techniques. ND, not detected. EB, embryoid body. (**c**) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3) and *NANOS2* (N2) for H1 in triplicate (columns) in 100, 50, 10, 0 or single TRA-1-81$^+$/cKIT$^+$ cells (rows) sorted from H1 and UCLA1 hESCs using five alternative differentiation strategies.

ilar to what is observed in cKIT$^+$ PGCs from the human fetus (Fig. 6e,f and Supplementary Fig. S3c). Immunofluorescence microscopy also revealed that all NANOS3/OCT4A$^+$ cells in the day-9 embryoid body were positive for 5mC (Fig. 6e). Taken together, our data suggest that TRA-1-81$^+$/cKIT$^+$/O/B/N3 cells in embryoid bodies correspond to human PGCs before gonadal colonization and loss of 5mC.

Using PGC differentiation in embryoid bodies for 9 days as a comparison, we show that O/B/N3 putative PGCs are transient, being lost by day 15 of embryoid body formation (Fig. 7b,c). Transferring day-9 embryoid bodies to hFGSCs for a further 7 days (16 days total) was consistent with an increase in survival and/or differentiation of O/B/N3 triple-positive cells; however, *DAZL* or *VASA* RNA was not induced in the O/B/N3 population (Fig. 7c). Sustaining the O/B/N3 triple-positive population within the TRA-1-81$^+$/cKIT$^+$ fraction for 15 days was also achieved using adherent differentiation on GFR M/G (Fig. 7c).

## DISCUSSION

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

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

A major event in reprogramming 2 is the erasure of cytosine methylation from DMRs of imprinted genes, which is thought to be active in mice[13]. In the present study we show that 5hmC and the *TET* enzymes are dynamically expressed by cKIT$^+$ human PGCs, as well as other molecular candidates that could actively modify 5mC/5hmC or remove modified 5hmC from the genome, including *AICDA* and *TDG* (refs 14,19–21,30,31). In hESCs and fibroblasts where cytosine methylation at imprinted DMRs is stably inherited, 5hmC is not detected at the *PEG3* DMR. In contrast in cKIT$^+$ PGCs, where the fate of this locus is demethylation, 5hmC is enriched. Despite this tantalizing

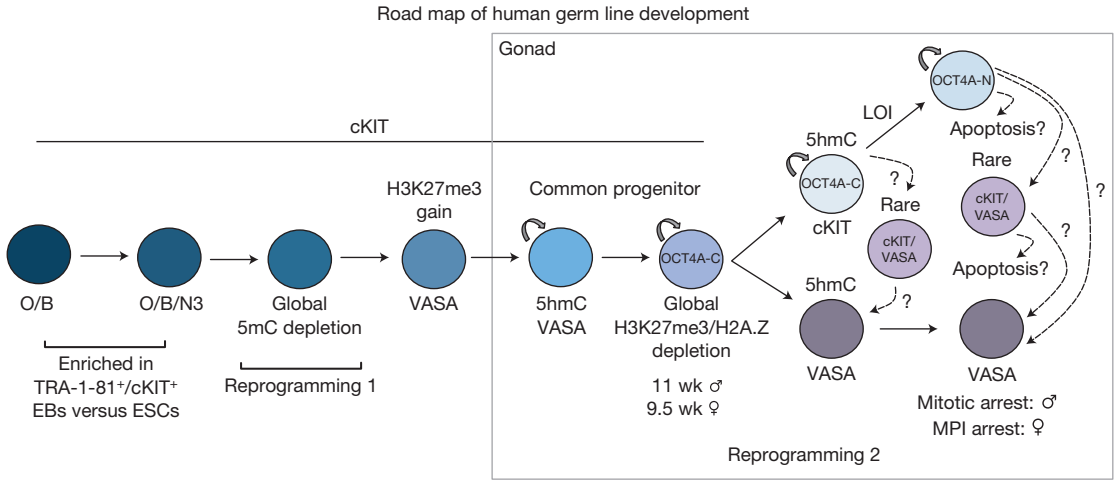


**Figure 8** Summarized roadmap of human germ line development. Reprogramming 1 occurs before 6–7 developmental weeks and is characterized by global loss of 5mC from PGC DNA. Reprogramming 2 begins in the common PGC progenitor stage after acquisition of H3K27me3 (10.5 weeks in testes and 8.5 weeks in ovaries) and involves global loss of H3K27me3 and H2A.Z followed by imprint erasure in cKIT$^+$ PGCs more than 1 month later. *In vitro* hESC differentiation using TRA-1-81$^+$/cKIT$^+$ sorting generates a rare cKIT$^+$ PGC population that is O/B double positive or O/B/N3 triple positive and corresponds to newly specified PGCs, before reprogramming 1. wk, week; O/B, *OCT4*/*BLIMP1*; O/B/N3, *OCT4*/*BLIMP1*/*NANOS3*; LOI, loss of imprinting; MPI, meiotic prophase I. EB, embryoid body.

correlation, future studies are needed to determine the role of TETs and 5hmC in imprint erasure. One observation from our reference map is that imprint erasure occurs over weeks and in a locus-specific manner. This relatively long window for imprint erasure in cKIT$^+$ PGCs stands in contrast to the mouse where erasure at many imprinted loci occurs within 24 h (ref. 32). Therefore, our data support the idea that removal of 5mC from imprinted DMRs in humans may involve diverse, locus-specific and time-dependent strategies.

Using undifferentiated hESCs we show that germ line genes are expressed in a stochastic manner in the undifferentiated state, with only rare undifferentiated hESCs expressing the major PGC determinant *BLIMP1* (ref. 26). However, with *in vitro* differentiation using cKIT with TRA-1-81 we enriched for BLIMP1 expression in cKIT$^+$ cells, called O/B cells that we speculate represent the first lineage-restricted PGCs equivalent to E6.25 in mice[33,34]. Differentiation also results in enrichment of O/B/N3 triple-positive cKIT$^+$ cells that we speculate represent the next stage in PGC development after O/B and before 5mC loss in reprogramming 1. O/B/N3 cells were never observed in single hESCs sorted for TRA-1-60 and were found in less than 5% of TRA-1-80$^+$/cKIT$^+$ sorted undifferentiated hESCs. The fact that at a single-cell level *VASA* and *DAZL* were never co-expressed with O/B/N3 in TRA-1-81$^+$/cKIT$^+$ cells with differentiation does not refute previous findings using *DAZL* and *VASA* as markers to define germ line identity[4,24,35–41]. On the contrary, we propose that sorting for TRA-1-81$^+$/cKIT$^+$ specifically enriches for newly specified germ line cells before *DAZL* and *VASA* expression and that acquisition of this immature cell can be achieved regardless of differentiation strategy.

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

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note: Supplementary Information is available in the online version of the paper*

### ACKNOWLEDGEMENTS

The authors would like to thank the UCLA Translational Pathology Core Laboratory and the UCLA Gene and Cellular Core Laboratory for some of the gonadal samples used in this study. We also thank J. Hargan-Calvopina, M. Oliveros-Etter and S. Diaz-Perez for critical reading of the manuscript, F. Codrea and J. Scholes for FACS and S. Peckman from the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research for critical assistance with human subject and embryonic stem cell review. This work was supported primarily by fund number 1R01HD058047 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD; ATC), as well as the Iris Cantor-UCLA Women's Health Pilot Project (ATC) and 1P01GM081621 from NIGMS. The Laboratory of Developmental Biology, University of Washington, Seattle is supported by NIH Award Number 5R24HD000836 from the NICHD. Human fetal tissue requests can be made to: bdrl@u.washington.edu.

### AUTHOR CONTRIBUTIONS

S.G. designed and performed the experiments, analysed data and wrote the manuscript; Z.L. performed flow analyses, RNA and DNA extraction for some gonadal samples, J.J.V. performed the single-cell analysis of female ovary at 14 weeks; K.X.Z. and M.P. performed the RNA-Seq data analysis; A.C. provided gonadal samples used in this study; A.T.C. designed the experiments, analysed data and wrote the manuscript.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at www.nature.com/doifinder/10.1038/ncb2638
Reprints and permissions information is available online at www.nature.com/reprints

1. Ohinata, Y. *et al.* A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137,** 571–584 (2009).
2. Hayashi, K. *et al.* Offspring from oocytes derived from *in vitro* primordial germ cell-like cells in mice. *Science* **16,** 971–975 (2012).
3. Reijo, R. *et al.* Mouse autosomal homolog of DAZ, a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty. *Genomics* **35,** 346–352 (1996).
4. Park, T. S. *et al.* Derivation of primordial germ cells from human embryonic and induced pluripotent stem cells is significantly improved by co-culture with human fetal gonadal cells. *Stem Cells* **27,** 783–795 (2009).
5. Gaskell, T. L., Esnal, A., Robinson, L. L. L., Anderson, R. A. & Saunders, P. T. K. Immunohistochemical profiling of germ cells within the human fetal testis: identification of three subpopulations. *Biol. Reprod.* **71,** 2012–2021 (2004).
6. Robinson, L., Gaskell, T., Saunders, P. & Anderson, R. Germ cell specific expression of c-kit in the human fetal gonad. *Mol. Human Reprod.* **7,** 845 (2001).
7. Høyer, P. E., Byskov, A. G. & Møllgård, K. Stem cell factor and c-Kit in human primordial germ cells and fetal ovaries. *Mol. Cell. Endocrinol.* **234,** 1–10 (2005).
8. Pauls, K. Spatial expression of germ cell markers during maturation of human fetal male gonads: an immunohistochemical study. *Hum. Reprod.* **21,** 397–404 (2005).
9. Kerr, C., Hill, C., Blumenthal, P. & Gearhart, J. Expression of pluripotent stem cell markers in the human fetal ovary. *Hum. Reprod.* **23,** 589–599 (2008).
10. Kerr, C. L., Hill, C. M., Blumenthal, P. D. & Gearhart, J. D. Expression of pluripotent stem cell markers in the human fetal testis. *Stem Cells* **26,** 412–421 (2008).
11. Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278,** 440–458 (2005).
12. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452,** 877–881 (2008).
13. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329,** 78–82 (2010).
14. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463,** 1101–1105 (2010).
15. Sinclair, A. H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346,** 240–244 (1990).
16. Atlasi, Y., Mowla, S. J., Ziaee, S. A. M., Gokhale, P. J. & Andrews, P. W. OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. *Stem Cells* **26,** 3068–3074 (2008).
17. Warthemann, R., Eildermann, K., Debowski, K. & Behr, R. False-positive antibody signals for the pluripotency factor OCT4A (POU5F1) in testis-derived cells may lead to erroneous data and misinterpretations. *Mol. Hum. Reprod.* **18,** 605–612 (2012).
18. Ruzov, A. *et al.* Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res.* **21,** 1332–1342 (2011).
19. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133 (2010).
20. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468,** 839–843 (2010).
21. Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Stem Cell* **8,** 200–213 (2011).
22. Thomson, J. A. Embryonic stem cell lines derived from human blastocysts. *Science* **282,** 1145–1147 (1998).
23. Diaz Perez, S. V. *et al.* Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression *in vitro* that can be prevented by chemical modification of chromatin. *Hum. Mol. Genet.* **21,** 751–764 (2012).
24. Clark, A. T. Spontaneous differentiation of germ cells from human embryonic stem cells *in vitro*. *Hum. Mol. Genet.* **13,** 727–739 (2004).
25. Zwaka, T. P. A germ cell origin of embryonic stem cells? *Development* **132,** 227–233 (2005).
26. Bao, S. *et al.* The germ cell determinant blimp1 is not required for derivation of pluripotent stem cells. *Cell Stem Cell* **11,** 110–117 (2012).
27. Kamei, K. *et al.* Microfluidic image cytometry for quantitative single-cell profiling of human pluripotent stem cells in chemically defined conditions. *Lab. Chip* **10,** 1113–1119 (2010).
28. Briddell, R. *et al.* Further phenotypic characterization and isolation of human hematopoietic progenitor cells using a monoclonal-antibody to the c-kit receptor. *Blood* **79,** 3159–3167 (1992).
29. Anderson, R., Fulton, N., Cowan, G., Coutts, S. & Saunders, P. Conserved and divergent patterns of expression of DAZL, VASA and OCT 4 in the germ cells of the human fetal ovary and testis. *BMC Dev. Biol.* **7,** 136 (2007).
30. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463,** 1042–1047 (2010).
31. Maiti, A. & Drohat, A. C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286,** 35334–35338 (2011).
32. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117,** 15–23 (2002).
33. Ohinata, Y., Sano, M., Shigeta, M., Yamanaka, K. & Saitou, M. A comprehensive, non-invasive visualization of primordial germ cell development in mice by the Prdm1-mVenus and Dppa3-ECFP double transgenic reporter. *Reproduction* **136,** 503–514 (2008).
34. Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22,** 1617–1635 (2008).
35. Kee, K., Gonsalves, J. M., Clark, A. T. & Pera, R. A. R. Bone morphogenetic proteins induce germ cell differentiation from human embryonic stem cells. *Stem. Cells Dev.* **15,** 831–837 (2006).
36. Kee, K., Angeles, V. T., Flores, M., Nguyen, H. N. & Pera, R. A. R. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature* **462,** 222–225 (2009).
37. Aflatoonian, B. *et al.* *In vitro* post-meiotic germ cell development from human embryonic stem cells. *Hum. Reprod.* **24,** 3150–3159 (2009).
38. Tilgner, K. *et al.* Expression of GFP under the control of the RNA helicase VASA permits fluorescence-activated cell sorting isolation of human primordial germ cells. *Stem Cells* **28,** 84–92 (2010).
39. Panula, S. *et al.* Human germ cell differentiation from fetal-and adult-derived induced pluripotent stem cells. *Hum. Mol. Genet.* **20,** 752–762 (2011).
40. Medrano, J. V., Ramathal, C., Nguyen, H. N., Simon, C. & Reijo Pera, R. A. Divergent RNA-binding proteins, DAZL and VASA, induce meiotic progression in human germ cells derived *in vitro*. *Stem Cells* **30,** 441–451 (2012).
41. Chuang, C. Y. *et al.* Meiotic Competent human germ cell-like cells derived from human embryonic stem cells induced by BMP4/WNT3A signaling and OCT4/EpCAM (epithelial cell adhesion molecule) selection. *J. Biol. Chem.* **287,** 14389–14401 (2012).
42. Plath, K. Role of histone H3 Lysine 27 methylation in X inactivation. *Science* **300,** 131–135 (2003).

## METHODS

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

**CGRA at the *PEG3* DMR.** For CGRA analysis at the *PEG3* DMR, DNA was extracted using the Quick-gDNA MiniPrep Kit (Zymo Research) according to the manufacturer's instructions. CGRA was next performed using the EpiMark 5hmC and 5mC Analysis Kit (NEB) according to the manufacturer's instructions followed by qRT–PCR for *PEG3* DMR DNA using primers PEG3 forward: 5′-CCACCTGCAGCCACTTC-3′ and PEG3 reverse: 5′-AGTTGGTTGGGCGAGACAAG-3′ with SYBR Green Master Mix (Applied Biosystems) at $T_m = 65\,^{\circ}C$ on a Bio-rad MyiQ Thermal Cycler (Bio-rad).

**Cell culture.** H1 (WA01, 0043, 46XY) and UCLA1 (0058, 46XX) hESC lines were maintained under self-renewal conditions on a mouse embryonic fibroblast layer in Dulbecco's Modified Eagle's Medium (DMEM)/F12 (Gibco BRL), 20% KnockOut Serum (Gibco BRL), 1% non-essential amino acids (NEAA, Gibco BRL), 1 mM L-glutamine (Gibco BRL), 0.1 mM β-mercaptoethanol (Gibco BRL) and 10 ng ml$^{-1}$ of basic fibroblast growth factor from the Biological Resources Branch of the Frederick National Laboratory for Cancer Research. Undifferentiated hESC colonies were passaged every 7 days and maintained as previously described[4]. For embryoid body differentiation, on day four after passage undifferentiated H1 and UCLA1 cells were treated with collagenase type IV (Gibco BRL; 1 mg ml$^{-1}$) for 45 min at 37 °C. The colonies that dissociated were lifted by gentle pipetting and plated overnight in ultralow attachment 6-well plates (Corning Incorporated) in mTeSR1 medium (StemCell Technologies) supplemented with 10 nM ROCK inhibitor (HA-1077, Sigma-Aldrich). At 24 h after initial plating, the medium was replaced by differentiation medium, DMEM/F12 supplemented with 20% FBS, 0.1 mM non-essential amino acids, 0.1 mM β-mercaptoethanol and 1 mM L-glutamine (all reagents from Gibco BRL). The medium was changed every third day. The addition of 50 ng ml$^{-1}$ carrier-free BMP4 (R&D Systems) was used where necessary. Adherent differentiation was performed on plates coated with GFR M/G (BD Pharmigen) in differentiation medium, changed every two days. For differentiation on FGSCs, differentiation medium was changed every two days. FGSCs were cultured as previously described[4]. For all experiments, hESCs were used between passages 8 and 45. All hESC experiments were conducted with prior approval from the UCLA Embryonic Stem Cell Research Oversight Committee. BJ fibroblast somatic cells were cultured in minimum essential medium with Earle's salt (Gibco BRL) and 1 mM L-glutamine, 10% FBS (Gibco BRL), 1% NEAA and 1 mM sodium pyruvate (Gibco BRL). HEK293T cells were maintained on gelatin-coated plates in DMEM High Glucose (Gibco BRL) supplemented with 1 mM L-glutamine, 10% FBS (Gibco BRL), 1% NEAA and 1 mM sodium pyruvate (Gibco BRL). BJ and HEK293T cells were passaged using 0.25% trypsin (Gibco BRL) every 5 days.

**FACS and flow cytometry.** Cells from embryoid bodies were dissociated in 0.25% trypsin–EDTA (Gibco BRL) supplemented with 2% chicken serum (Gibco BRL) at 37° C for 30 min and collected by centrifugation at 200*g* in an Eppendorf 5702 R centrifuge fitted with an A-4-38 swing bucket rotor for 5 min. For adherent differentiation, human fetal gonads and HEK293T, cells were dissociated with 0.25% Trypsin–EDTA for 5 min at 37 °C. Dissociated cells were incubated in 1% BSA in PBS containing primary antibodies on ice for 20 min. Primary antibodies used for flow analyses were: cKIT–APC conjugated (1:100, 550412; BD Biosciences), TRA-1-81 (1:100, 14-8883; eBioscience), TRA-1-60 (1:100, 14-8863; eBioscience) and CD45-PE conjugated (1:50, PN IM1833; Beckman-Coulter). Cells were then washed and incubated with FITC-conjugated (Jackson ImmunoResearch) or Pe–Cy7-conjugated secondary antibody (BD Biosciences) on ice for another 20 min. Cells were washed again and incubated with 1% BSA in PBS for 5 min on ice. To ensure single-cell separation before flow analysis or FACS sorting, cells were passed through a 40 μM filter (BD Biosciences). As a viability dye, for *in vitro* differentiation 7-AAD (1:50, BD Pharmigen) was used, and for gonadal samples DAPI at 1:1,000 dilution (10 μg ml$^{-1}$, Sigma) was used. Analysis was performed using LSR II (Becton Dickinson) and FlowJo software (Tree Star).

**Immunofluorescence microscopy.** Immunofluorescence microscopy of fetal gonads from 6 to 19 weeks of gestation was performed as previously described[4]. Dilutions and catalogue numbers of primary antibodies used were: cKIT (1:100, A4502; DAKO), OCT4A (1:100, (N-19)-sc-8628; Santa Cruz), VASA (1:400, AF2030; R&D Systems), H3K27me3 (1:800, 05-1339; Millipore), H2A.Z (1;100, ab4174; Abcam); 5mC (1:100, AMM99021; Aviva), NANOS3 (1:300, ab70001; Abcam), SYCP3 (1:50, NB300-232; Novus Biologicals), BLIMP1 (1:100, 9115; Cell Signaling), TRA-1-81 (1:100; 14-8883; eBioscience), SSEA1 (1:50, MC-480; Jackson ImmunoResearch) and 5hmC (1:100, 39769; Active Motif). For 5hmC, a denaturing step was added to the staining procedure as previously described[43]. All samples were incubated with primary antibodies overnight at 4 °C. Sections were washed, incubated with FITC/TRITC-conjugated secondary antibodies (Jackson ImmunoResearch) for 30 min and mounted in Prolong Antifade Reagent with DAPI (Invitrogen). Samples were imaged on a Zeiss Axio Imager (Zeiss) using Axio Vision 4.8 Software (Zeiss).

**Single-cell real-time RT–PCR.** Single cells were sorted with a BD ARIA cell sorter equipped for Biosafety Level 2 (BSL2) sorting, and single-cell analysis was performed as described previously[44].

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

**RNA-Seq analysis.** We performed single-end RNA-Seq analysis and obtained 116,399,186 (UW30) and 153,982,474 (UW99) total reads, of which 35,497,900 and 37,907,565, respectively, were uniquely mapped to the reference genome (UCSC assembly hg19) for the cKIT$^+$ fetal testis libraries. For the cKIT$^+$ fetal ovary libraries, we obtained 211,659,065 (UW92) and 102,837,931 (UW98) total reads, of which 93,766,963 and 44,784,216 reads, respectively, were uniquely mapped. For the three H1 hESC replicates we obtained 161,037,803 (A), 133,355,463 (B) and 104,997,235 (C) total reads, of which 39,339,764, 38,571,046 and 37,164,448 reads, respectively, were uniquely mapped. Reads were mapped to the reference genome using the gapped aligner Tophat (version 1.3.0) allowing up to 2 mismatches. The human gene model annotation (version of Homo_sapiens.GRCh37.59) was downloaded from the Ensembl database and supplied to Tophat. Gene and transcript expression levels were quantified using Cufflinks (version 1.0.3) in the FPKM unit (fragments per kilobase of exon per million fragments mapped) together with confidence intervals. Cufflinks ran in the default parameters except that the annotated gene set was supplied using the -G option. The raw read count for each gene and transcript was measured using customized scripts written in Perl. RNA Sequencing data are available in GEO with accession number GSE39821.

**Differential expression testing.** Differential expression analysis was performed using the R package, DESeq and edgeR. Raw read counts were employed and modelled as negative binomial distributed because our data set contains biological replicates. We filtered out lowly expressed genes and transcripts by keeping only those that have at least one count per million in samples. The multiple testing errors were corrected by the false discovery rate. Besides the cutoff as adjusted at $P < 0.05$, we also adopted another cutoff set as the expression ratio of above twofold changes in expression values. In summary, we considered genes as differentially expressed if: the adjusted $P$ value was less than 0.05; the expression ratio between two conditions was above twofold; there was agreement between DESeq and edgeR. The heat map in Fig. 5c was generated using the default settings with the function 'heatmap.2' in the 'gplots' package of R. RPKM values are centred and scaled by subtracting the mean of the row from every RPKM value and then dividing the resulting RPKM values by the standard deviation of the row.

DOI: 10.1038/ncb2638

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

**Statistical analyses and graphs.** All data are expressed as mean ± s.e.m. and statistical analyses were performed using the Mann–Whitney test. Graphs were generated using GraphPad Prism 5 (GraphPad Software).


43. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).
44. Vincent, J. J. *et al.* Single cell analysis facilitates staging of blimp1-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS ONE* **6,** e28960 (2011).
45. Boissonnas, C. *et al.* Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eue. J. Human Gene.* **18,** 73–80 (2009).
46. Geuns, E., Hilven, P., Van Steirteghem, A., Liebaers, I. & De Rycke, M. Methylation analysis of KvDMR1 in human oocytes. *J. Med. Genet.* **44,** 144–147 (2006).
47. Kagami, M. *et al.* The IG-DMR and the MEG3-DMR at human chromosome 14q32.2: hierarchical interaction and distinct functional properties as imprinting control centers. *PLoS Genet.* **6,** e1000992 (2010).
48. Zechner, U. *et al.* Quantitative methylation analysis of developmentally important genes in human pregnancy losses after ART and spontaneous conception. *Mol. Hum. Reprod.* **16,** 704–713 (2010).

nature cell biology

DOI: 10.1038/ncb2638

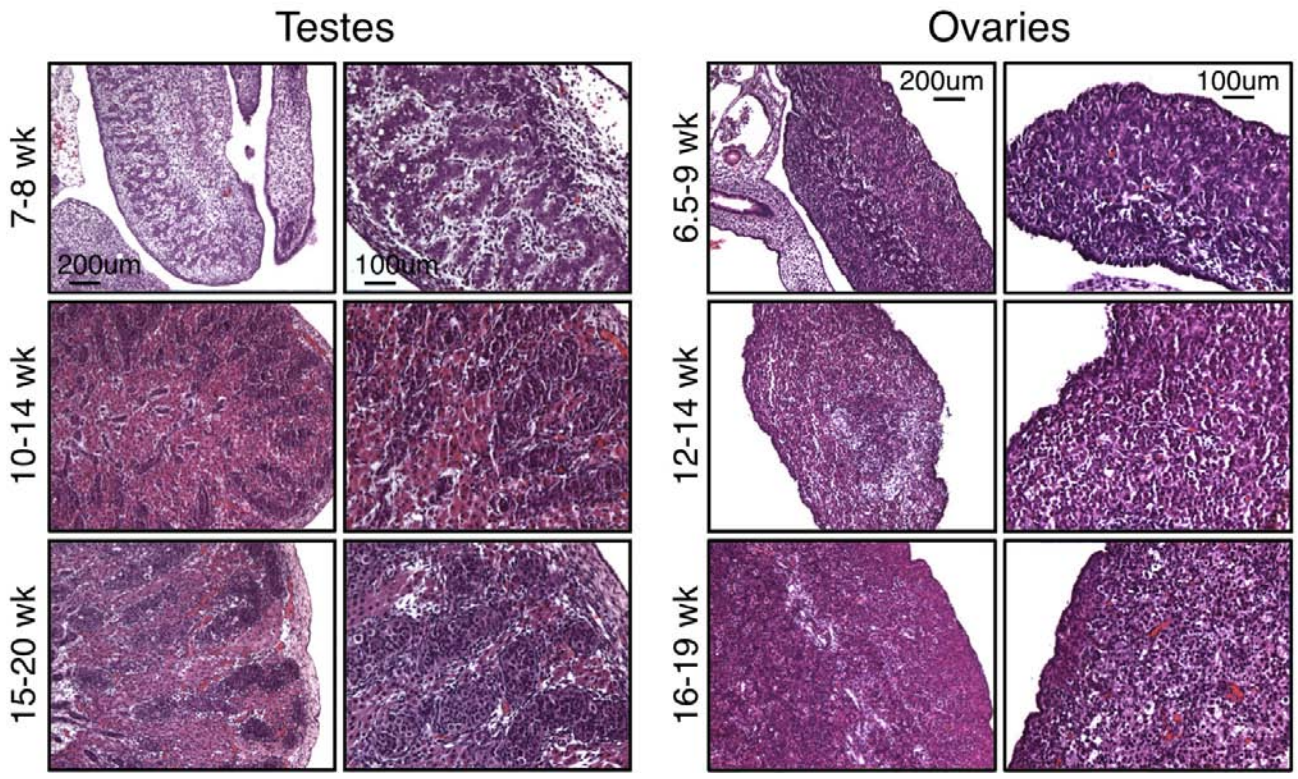


**Figure S1** Morphological characteristics of fetal testes and ovaries from 6.5-20 developmental weeks. Representative images of Hematoxylin and Eosin staining of testes and ovaries over the developmental time period analyzed in this project. In testes, shown is an 8-week for 7-8wk (n=2), a 10-week for 10-14wk (n=3) and a 15-week for 15-20wk (n=5). In ovaries, shown at 6.5-week for 6.5-9wk (n=4), a 14-week for 12-14wk (n=2) and an 18-week for 16-19wk (n=3).

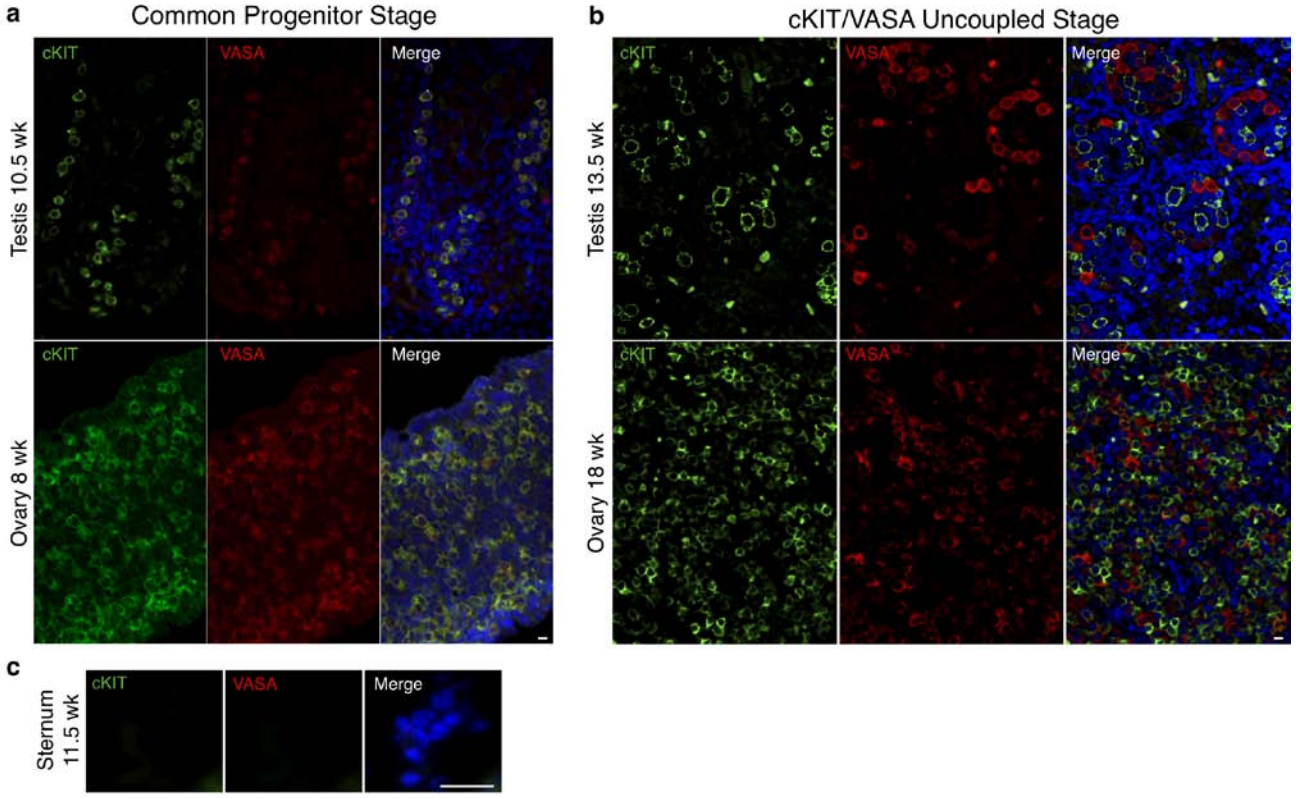


**Figure S2** cKIT and VASA expression in fetal testes and ovaries. Representative immunofluorescence images of cKIT with VASA, **(a)** at the common progenitor stage, in a 10.5-week testes and an 8-wk ovary, (**b**) at the cKIT/VASA uncoupled stage, in a 13.5-week testis and an 18-week ovary. **(c)** Immunofluorescence of cKIT with VASA on 11.5-week sternum tissue as negative control for cKIT and VASA antibody specificity on paraffin sections. Nuclei were counterstained with DAPI (blue). Scale bars represent 10 um. Abbreviations: wk= week.

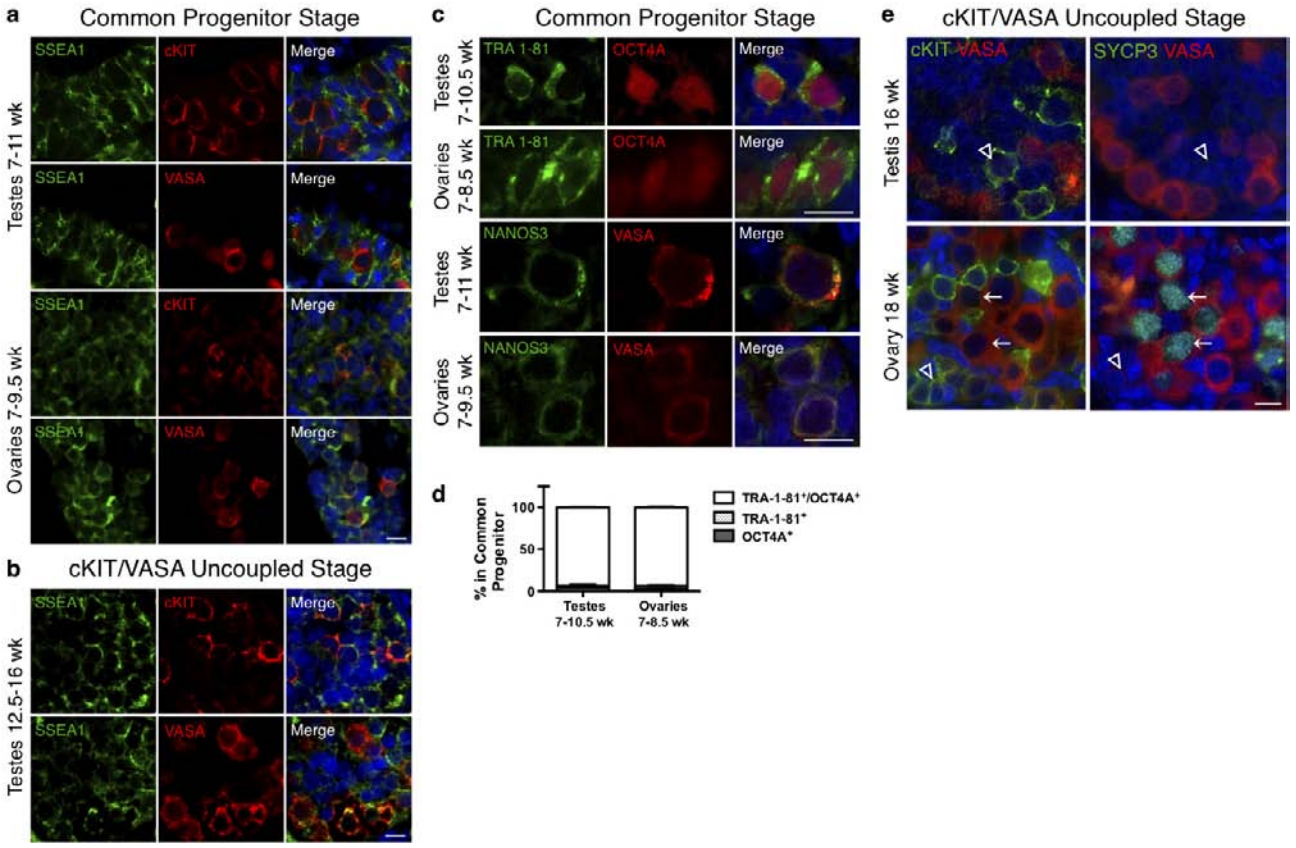


**Figure S3** Dynamics of cKIT, OCT4 and/or VASA with SSEA1, TRA-1-81, NANOS3 and SYCP3. (**a,b**) Representative immunofluorescence images of SSEA1 with cKIT or VASA, (**a**) at the common PGC progenitor stage, in testes shown is an 11-week sample for 7-11wk (n=2) and an 8-week ovary for 7-9.5wk (n=2). (**b**) The cKIT/VASA uncoupled stage in the testis from 12.5-16wk (n=3) shown is a testes at 13.5-weeks. (**c**) Representative immunofluorescence images of TRA-1-81 with OCT4A and NANOS3 with VASA at the common PGC progenitor stage. For TRA-1-81, shown is a 10.5-week testis for 7-10.5wk (n=3) and an 8.5-week ovary for 7-8.5wk (n=3). For NANOS3 shown is an 11-week in testis for 7-11 wk (n=3) and a 9.5-week in ovary for 7-9.5wk (n=2). (**d**) Quantification of TRA-1-81$^+$, OCT4A$^+$ and TRA-1-81$^+$/OCT4A$^+$ in testes at 7-10.5wk, 6 optic fields counted (n=3) and in ovaries at 7-8.5wk, 7 optic fields counted (n=3). (**e**) Representative immunofluorescence images at the cKIT/VASA uncoupled stage of VASA with cKIT or SYCP3 on adjacent sections at 16 weeks in testis and 18 weeks in ovary. Arrows indicate the same cKIT$^+$, and open arrowheads the same VASA$^+$ cell in adjacent sections. Nuclei were counterstained with DAPI (blue). Scale bars represent 10 um. All data are mean ± SEM.

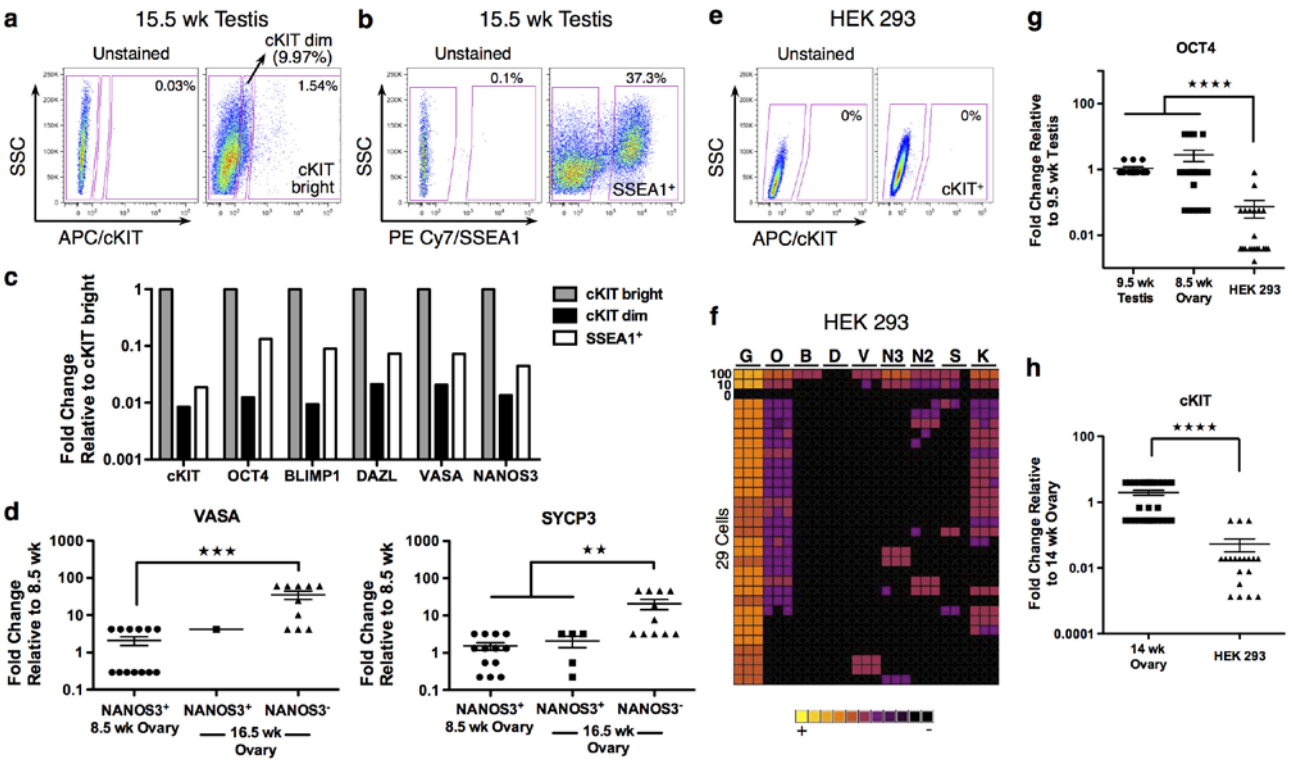


**Figure S4** Germ line identity is enriched in the cKIT bright fraction in 15.5-week testis. (**a**,**b**) Gating strategy for sorting, (**a**) cKIT dim and cKIT bright cells with an APC conjugated anti-human cKIT primary antibody and (**b**) SSEA1$^+$ cells using PE Cy7 secondary antibody against side scatter (SSC) on a 15.5-week testis. Indicated above the graph is the percent positive cells sorted in each gate (**c**) qRT-PCR showing fold change enrichment in cKIT dim and SSEA1$^+$ relative to the cKIT bright fraction for *cKIT*, *OCT4*, *BLIMP1*, *DAZL*, *VASA* and *NANOS3* after *GAPDH* normalization. Shown on a log10 scale. (**d**) qRT-PCR data analysis for *VASA* and *SYCP3* in *NANOS3* expressing cKIT$^+$ cells sorted from the ovary at 8.5 weeks, and *NANOS3* positive or negative cKIT$^+$ cells sorted from the ovary at 16.5 weeks. Levels are normalized to *GAPDH* and are shown as fold change relative to 8.5 weeks on a log10 scale. Each data point represents n=1 single cell. (**e**) Flow plot of HEK 293 stained with an APC conjugated anti-human cKIT primary antibody against side scatter (SSC), showing there are no cKIT expressing cells. (**f**) Heat map of *GAPDH* (G), *OCT4* (O), *BLIMP1* (B), *DAZL* (D), *VASA* (V), *NANOS3* (N3), *NANOS2* (N2), *SYCP3* (S) and *cKIT* (K) in triplicate (columns) in 100, 10, 0 or single cells (rows) in sorted HEK 293 cells. Note that HEK 293 cells have previously been reported to express *OCT4* mRNA[15] and we show that cKIT is transcribed but the protein is not expressed on the cell surface (**g**) qRT-PCR analysis for *OCT4* in single cKIT$^+$ cells sorted from the fetal testes at 9.5 weeks, the fetal ovary at 8.5 weeks, and sorted HEK 293 cells. (**h**) qRT-PCR data analysis for *cKIT* in cKIT$^+$ cells sorted from the fetal ovary at 14 weeks and sorted HEK 293s. Levels are normalized to *GAPDH* and are shown as fold change referenced to the 9.5 wk testis in (**g**) and the 14 wk ovary in (**h**) on a log10 scale. Note that where individual HEK 293 cells express *OCT4* and *cKIT* the levels are 10-to-100 fold lower than the reference germ line cells. Each data point represents n=1 cell. All data are mean ± SEM. (*P* values were determined using a Mann-Whitney test, ⋆⋆ p<0.005, ⋆⋆⋆ p<0.0005 and ⋆⋆⋆⋆ p<0.0001).

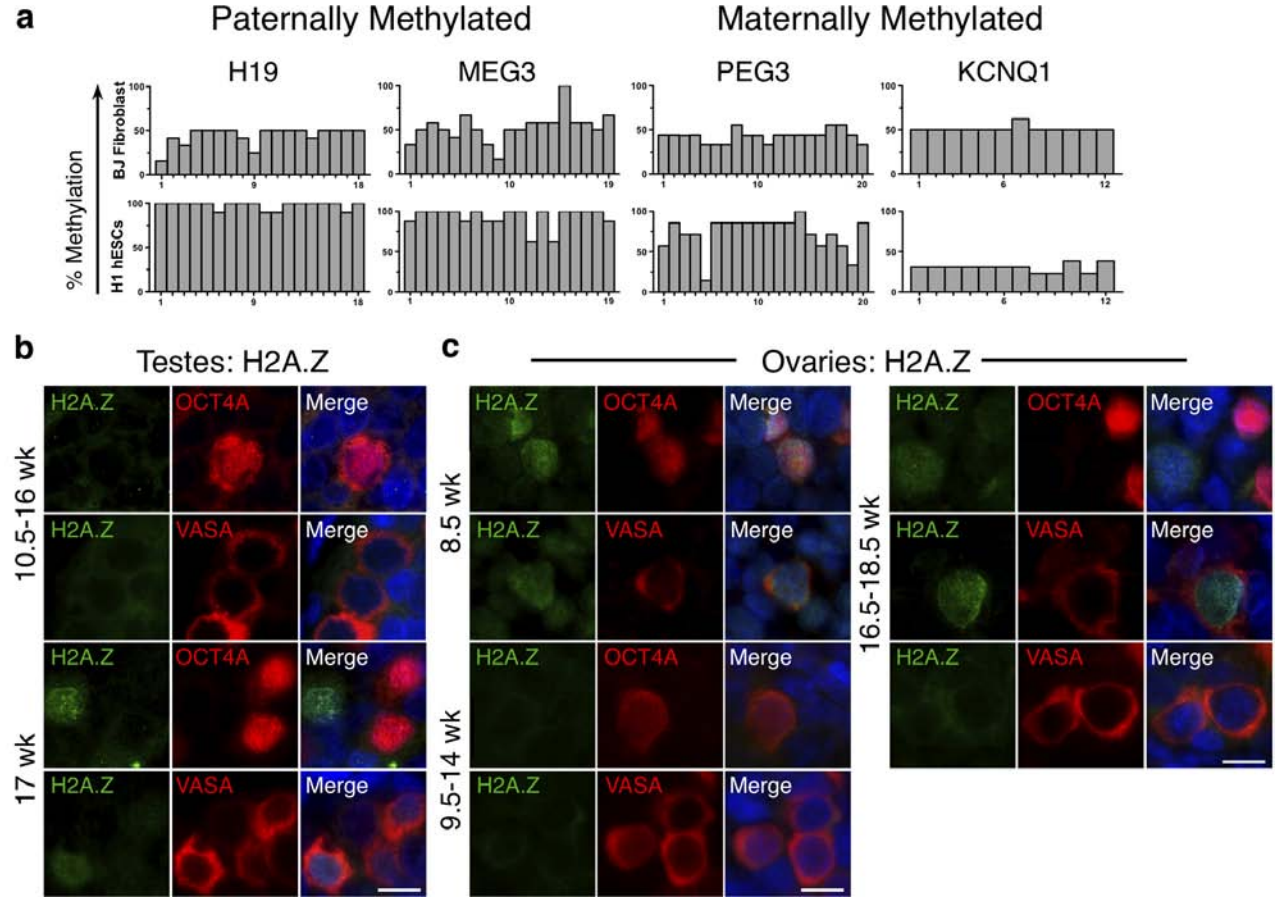


**Figure S5** (**a**) Methylation Analysis of BJ fibroblast line and H1 hESC line performed by BS-PCR on *H19*, *MEG3*, *PEG3* and *KCNQ1* DMRs in BJ fibroblast and H1 hESCs. (**b,c**) Fetal PGCs in the testis and ovary become transiently devoid of H2A.Z. Representative immunofluorescence images of H2A.Z with OCT4A or VASA in (**b**) testes and (**c**) ovaries at designated developmental ages. (**b**) Shown is a 13.5-week testis for 10.5-16wk (n=5) and a 17-week testis (n=1). (**c**) Shown is an 8.5-week ovary (n=2), a 14-week ovary for 9.5-14wk (n=3) and a 18.5-week ovary for 16.5-18.5wk (n=3). Abbreviations, wk=week. Nuclei were counterstained with DAPI (blue). Scale bars represent 10 um.

**Table S1**

| | Assay \ Age | 6.5-10.5wk | 10.6-12.5wk | 12.6-14.5wk | 14.6-16.5wk | 16.6-20wk |
|---|---|---|---|---|---|---|
| Male n=73 | FACS | 7 | 9 | 14 | 15 | 4 |
| | IF | 6 | 5 | 7 | 5 | 1 |
| Female n=61 | FACS | 15 | 6 | 12 | 8 | 1 |
| | IF | 12 | 1 | 3 | 1 | 2 |