nature cell biology

Article

https://doi.org/10.1038/s41556-023-01224-7

DMRT1 regulates human germline commitment

Received: 23 February 2022

Accepted: 7 August 2023

Published online: 14 September 2023

Check for updates

Naoko Irie[1,2,11,12] ✉, Sun-Min Lee[1,3,11], Valentina Lorenzi[4,5], Haiqi Xu[6], Jinfeng Chen[6], Masato Inoue[6], Toshihiro Kobayashi[7,8], Carmen Sancho-Serra[4], Elena Drousioti[1], Sabine Dietmann[9], Roser Vento-Tormo[4], Chun-Xiao Song[6] & M. Azim Surani[1,10,12] ✉

Germline commitment following primordial germ cell (PGC) specification during early human development establishes an epigenetic programme and competence for gametogenesis. Here we follow the progression of nascent PGC-like cells derived from human embryonic stem cells in vitro. We show that switching from BMP signalling for PGC specification to Activin A and retinoic acid resulted in DMRT1 and CDH5 expression, the indicators of migratory PGCs in vivo. Moreover, the induction of DMRT1 and SOX17 in PGC-like cells promoted epigenetic resetting with striking global enrichment of 5-hydroxymethylcytosine and locus-specific loss of 5-methylcytosine at DMRT1 binding sites and the expression of DAZL representing DNA methylation-sensitive genes, a hallmark of the germline commitment programme. We provide insight into the unique role of DMRT1 in germline development for advances in human germ cell biology and in vitro gametogenesis.

Germ cells generate a totipotent zygote state at fertilization and transmit genetic and epigenetic information for development to term[1]. In humans, primordial germ cells (PGCs), the precursors of eggs and sperm, appear on approximately week 2 in gastrulating embryos[2–4]. The subsequent migration of PGCs into gonads over approximately weeks 5 to 6 is accompanied by critical epigenetic resetting[5–10]. A prolonged development and dormancy follow before functionally mature sperm and eggs form at puberty[2–4].

Technical and ethical reasons hamper research on early human PGCs, but in vitro models using human pluripotent stem (PS) cells, embryonic stem (ES) cells or induced PS cells have allowed advances in human germline biology, including the mechanism of PGC specification[4,11–15]. PS cells gain competence for germline fate following culture with GSK3 inhibitor and Activin A (ActA) called precursor of mesendoderm (preME)/incipient mesoderm-like cells (iMeLCs)[11,13], or with four inhibitors for GSK3, MEK, p38, JNK with FGF2, TGFβ and LIF (henceforth called 4i ES cells)[14], that differentiate into PGC-like cells (PGCLCs) in response to BMP2/BMP4. PGCLCs can also be induced directly in response to the ectopic expression of SOX17 and PRDM1, the essential regulators of human PGC fate[13,14]. The molecular programme of PGC specification is conserved in mammals with bilaminar disc embryos[13,16–22].

Development beyond the nascent stage has been explored with PGCs from aborted foetuses[5–10], but infrequently from week

[1]Wellcome Trust/Cancer Research UK Gurdon Institute, Henry Wellcome Building of Cancer and Developmental Biology, Cambridge, UK. [2]Metabolic Systems Laboratory, Live Imaging Center, Central Institute for Experimental Animals, Kanagawa, Japan. [3]Department of Physics, Konkuk University, Seoul, Republic of Korea. [4]Wellcome Sanger Institute, Cambridge, UK. [5]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK. [6]Ludwig Institute for Cancer Research and Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [7]Division of Mammalian Embryology, Center for Stem Cell Biology and Regenerative Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [8]Center for Genetic Analysis of Behavior, National Institute for Physiological Sciences, Aichi, Japan. [9]Department of Developmental Biology and Institute for Informatics, Washington University School of Medicine, St. Louis, MO, USA. [10]Physiology, Development and Neuroscience Department, University of Cambridge, Cambridge, UK. [11]These authors contributed equally: Naoko Irie, Sun-Min Lee. [12]These authors jointly supervised this work: Naoko Irie, M. Azim Surani. ✉e-mail: irienaoko@ciea.or.jp; a.surani@gurdon.cam.ac.uk

2 to week 6, because of their scarcity. The nascent PGCs in vivo originating at gastrulation, proliferate and migrate through the yolk sac endoderm, hindgut and dorsal mesentery before they reach developing gonads from approximately week 5 onwards[2–4]. Critical epigenetic resetting accompanies PGC migration, including a transient enrichment for DNA 5-hydroxymethylcytosine (5hmC) and progressive global loss of 5-methylcytosine (5mC) that diminishes to ~5% by week 9 (refs. 6–9). The expression of *DAZL*, *PIWIL1* and *PIWIL2* in PGCs as they enter gonads marks lineage commitment, while the expression of nascent PGC genes, *SOX17*, *PRDM1*, *NANOS3*, *POU5F1* and *NANOG* continues[6,23,24]. With germline commitment, there is suppression of pluripotency genes after proliferation when female PGCs commence meiosis followed by oogenesis at approximately weeks 11 to 14, and male PGCs undergo mitotic arrest at approximately week 9 (ref. 24). Failure of the commitment may lead to germ cell tumourigenesis[25,26].

Nascent PGCLCs induced from PS cells exhibit the potential to develop further, albeit at a low frequency, when co-cultured with hindgut organoids[27] or mouse gonadal somatic cells[28,29]. However, the underlying mechanisms of development, epigenetic resetting, global DNA demethylation and migratory/gonadal PGC gene expression, including DAZL, remain unclear.

In this Article, we report the development of PGCLCs beyond the nascent stage in defined conditions and reveal the hitherto unknown role of DMRT1 in the transition from nascent PGCs towards germline commitment. DMRT1, with a DM domain and zinc finger-like DNA binding motif, is evolutionarily conserved and is best known for its role in sex determination and spermatogenesis[30–33]. Our study provides mechanistic insights into the role of DMRT1 in the stepwise developmental progression of human germ cell lineage.

Results

Sustained signalling for PGCLCs restricts progression

First, we compared the transcriptome of nascent PGCLCs (equivalent to weeks 2 to 3 PGCs in vivo) with the in vivo tissue non-specific alkaline phosphatase (TNAP)$^+$KIT$^+$ PGCs from week 5, week 7 and week 9 foetuses[6,14]. We classified PGC(LCs) as migratory, mitotic and mitotic arrest (gonocytes) as described previously (Methods)[24]. Nascent PGCLCs showed *SOX17*, *PRDM1* and *NANOS3* expression but without *CDH5* or *DMRT1*, which occurs in migratory PGCs of approximately week 4 (Fig. 1a and Extended Data Fig. 1a)[24]. We detected CDH5 in male and female gonadal PGCs (Fig. 1b and Extended Data Fig. 1b). Progressive upregulation of *DMRT1* in the mitotic and mitotic arrest PGCs was found, which was followed by *DAZL* expression, including subpopulation of migratory PGCs at week 5 (ref. 27), indicating a step-wise gene expression for germline commitment (Fig. 1a).

To monitor the progression of nascent PGCLCs, we established male and female ES cells with DMRT1 and NANOS3 dual fluorescent reporters (Fig. 1c and Extended Data Fig. 1c,d)[13,14]. These cells also contained dexamethasone (DEX)-inducible SOX17 and doxycycline (dox)-inducible PRDM1 factor for PGCLC induction without the cytokines (Fig. 1c)[13], as well as anti-apoptotic BCL2L1.

The NANOS3–tdTomato-positive PGCLCs were induced in response to BMP2, SCF and EGF (BSE), or SOX17/PRDM1, with or without BSE after 5 days (Fig. 1d). Unexpectedly, SOX17/PRDM1 also robustly induced DMRT1 (55–57.1%) in the PGCLCs in the absence of BSE (Fig. 1d and Extended Data Fig. 1e–g). With the optimal concentrations of DEX and dox for SOX17 and PRDM1 induction, respectively, dose-dependent regulation of PRDM1 by dox affected the induction of DMRT1. However, with the lower concentration of DEX regulating SOX17, there was efficient induction of DMRT1, suggesting that appropriate levels of the two factors are essential for further development (Extended Data Fig. 1i–k). Quantitative reverse transcription polymerase chain reaction (RT–qPCR) confirmed *DMRT1* expression in response to SOX17 and PRDM1 without BSE in NANOS3$^+$CD38$^+$ PGCLCs, but *CDH5* expression was not detectable (Fig. 1e and Extended Data Fig. 1f), unlike in migratory PGCs (Fig. 1a,b). However, retinoic acid (Ra) induced CDH5 in a dose-dependent manner in PGCLCs generated in response to SOX17/PRDM1, but without the expression of DMRT1 reporter (Fig. 1f). Conversely, ActA induced DMRT1 in PGCLCs without CDH5 (Fig. 1g,h). Notably, the combination of ActA and Ra resulted in robust expression of both DMRT1 (~66%) and CDH5 (~43%) in PGCLCs (Fig. 1h), indicating that they act independently to induce expression of DMRT1 and CDH5 in nascent PGCLCs to promote a migratory state.

Switching signalling promotes PGCLC progression

Since BSE induces PGCLC specification but hinders their subsequent development (Fig. 1d), we first induced PGCLCs with BSE and replaced it with Ra on day 2 and with Ra/ActA on day 3, which resulted in a robust expression of CDH5 and DMRT1 in PGCLCs over 8 days (38.6% NANOS3$^+$DMRT1$^+$ cells) (Extended Data Fig. 2a). The inclusion of SCF and EGF (SE), the PGCLC survival factors[34–36], together with Ra and ActA, enhanced the induction of DMRT1 and CDH5 (86.9% NANOS3$^+$DMRT1$^+$ cells on day 8) (Extended Data Fig. 2a). Indeed, the simultaneous addition of Ra and ActA with SCF and EGF (henceforth called RASE) when replacing BSE was most effective for the induction of DMRT1 (Extended Data Fig. 2b). ActA family members, TGF-β and Nodal that also activate SMAD2/3 did not induce DMRT1 (Extended Data Fig. 2c).

Nascent PGCLCs induced by BSE were responsive to RASE for CDH5/DMRT1 expression (henceforth DM$^+$PGCLCs; Fig. 8b) from day 2 onwards but not on day 1 (Fig. 2a–c), which also occurred in PGCLCs induced from female 4i ES cells and from preME/iMeLCs[11,13] (Extended Data Fig. 2d,e). DMRT1 protein co-localized with POU5F1 (Fig. 2d) and *SOX17*, *PRDM1*, *DMRT1* and *CDH5* transcripts were detected in PGCLCs induced by BSE followed by RASE (Fig. 2e).

We analysed the available single-cell RNA sequencing (scRNA-seq) of PGCs and gonadal soma[24] and found expression of ActA (encoded by *INHBA*) in the soma and of the receptors in PGCs (Fig. 2f).

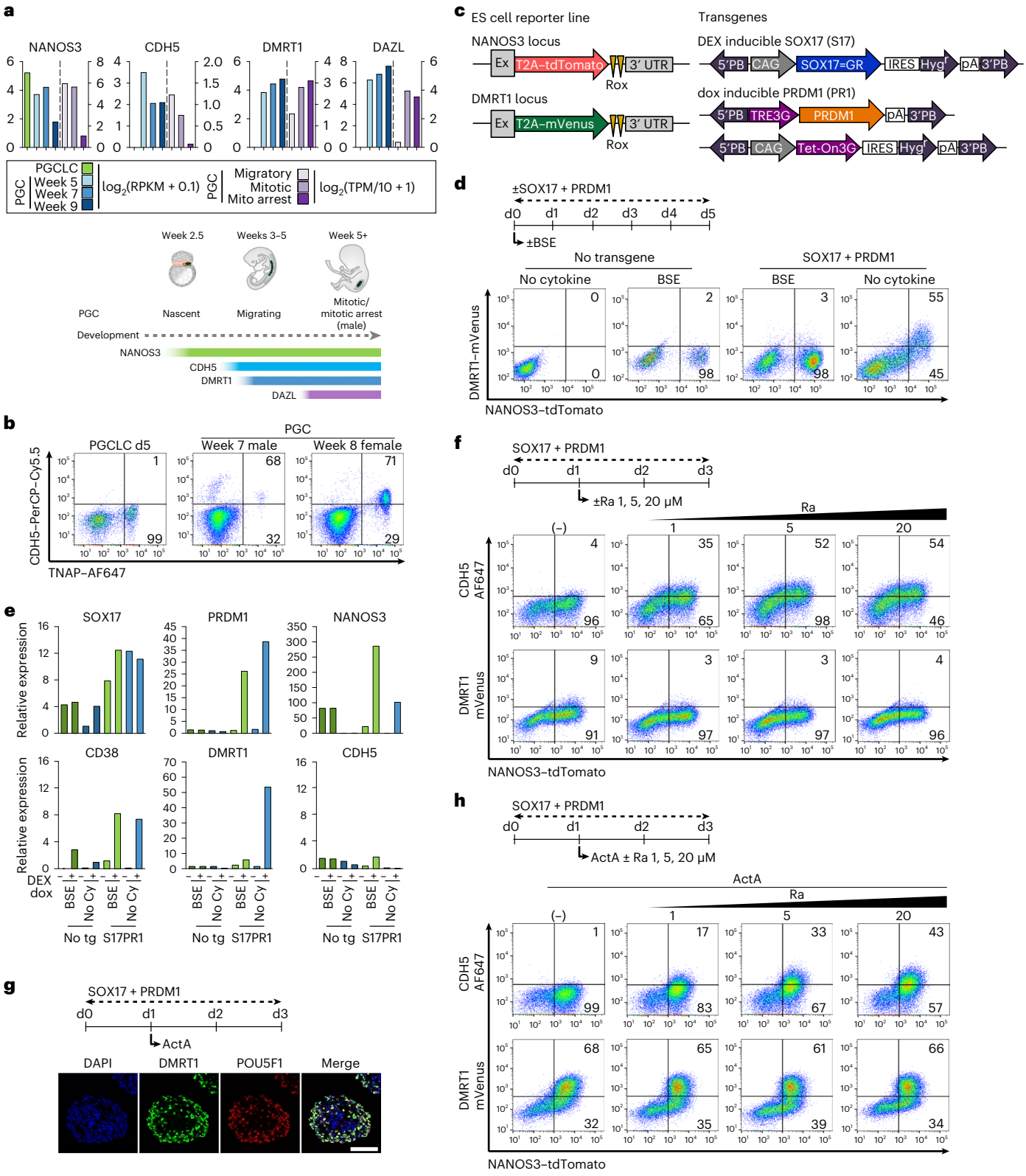
Fig. 1 | Sustained signalling for PGCLC specification restricts subsequent development. a, RNA-seq data[6,14,24] of in vitro induced day 4 PGCLCs and in vivo PGCs. Top: bar plot represents mean (PGC weeks 5/7/9, $n = 2$ biological samples; migratory, 37 cells; mitotic, 332 cells; mitotic arrest, 309 cells). Bottom: stage-specific gene expression during human PGC development. Vertical dashed lines separate the two datasets based on their different vertical scales. **b**, Flow cytometry of CDH5 and TNAP for in vitro induced day (d)5 PGCLCs and in vivo PGCs from week 7 male and week 8 female. Values are percentage ratio of CDH5-positive and CDH5-negative population in TNAP-expressing cells. **c**, PiggyBac (PB) gene introduction for DEX-inducible SOX17=GR and dox-inducible PRDM1 in NANOS3–tdTomato/DMRT1–mVenus double reporter ES cells. **d**,**e**, Flow cytometry (**d**) and RT–qPCR (**e**) of PGCLC induction from parental NANOS3–tdTomato/DMRT1-mVenus reporter WIS2 ES cells (no transgene, no tg) and SOX17=GR/TRE–PRDM1 cell line (S17PR1) clone 1 treated with (SOX17 + PRDM1, +) or without (−) DEX and dox for 5 days with BSE (BMP2, SCF and EGF) or without cytokines (no cytokine, no Cy). Values in **d** are percentage ratio of DMRT1–mVenus-positive and DMRT1–mVenus-negative population in NANOS3–tdTomato-expressing cells. Values in **e** are normalized with GAPDH and relative changes against no tg/no Cy/DEXdox (−). A repeat experiment with independent clone with similar results shown in Extended Data Fig. 1f. **f**, Flow cytometry of NANOS3–tdTomato against DMRT1–mVenus and CDH5–AF647. Induction of SOX17 + PRDM1 for 3 days, with (1, 5, 20 μM) or without (−) Ra from day 2. Percentage ratio of CDH5 and DMRT1 in NANOS3–tdTomato-positive cells. **g**, Immunofluorescence for DMRT1 and POU5F1 co-staining on day 3 of SOX17 + PRDM1 induction with ActA 100 ng ml^{-1} from day 1. Scale bar, 100 μm. The experiment was repeated independently two times with similar results. **h**, Flow cytometry of NANOS3–tdTomato against DMRT1–mVenus and CDH5–AF647. Induction of SOX17 + PRDM1 for 3 days with ActA 100 ng ml^{-1} from day 1 with (Ra; 1, 5, 20 μM) or without (−) Ra from day 1. Percentage ratio of CDH5 and DMRT1 in NANOS3–tdTomato-positive cells.

https://doi.org/10.1038/s41556-023-01224-7

Contemporarily, expression of *STRA6*, retinol receptor, and enzymes for Ra synthesis, *RDH11*, *ALDH1A1* and *ALDH1A3*, was present in soma (Fig. 2g). Ra receptor *RAR* family and *FABP5*, which delivers Ra to RARs[37,38], were expressed in migratory and mitotic PGCs, while *CYP26* enzymes for Ra degradation were low (Fig. 2g). Accordingly, PGCs in vivo can respond to ActA and Ra expressed in surrounding cells that can induce *DMRT1* and *CDH5*.

DMRT1 can activate DAZL

Expression of DAZL, a DNA methylation-sensitive PGC gene, accompanies a decline in 5mC levels in PGCs, indicating germline commitment[39–42]. A lack of DAZL expression in DM$^+$PGCLCs indicates that they are at an earlier stage with substantial overall levels of 5mC (Fig. 2e and Extended Data Fig. 2f). To investigate further, we established ES cell lines with a DAZL reporter (Extended Data Figs. 1c,d and 3b).



https://doi.org/10.1038/s41556-023-01224-7

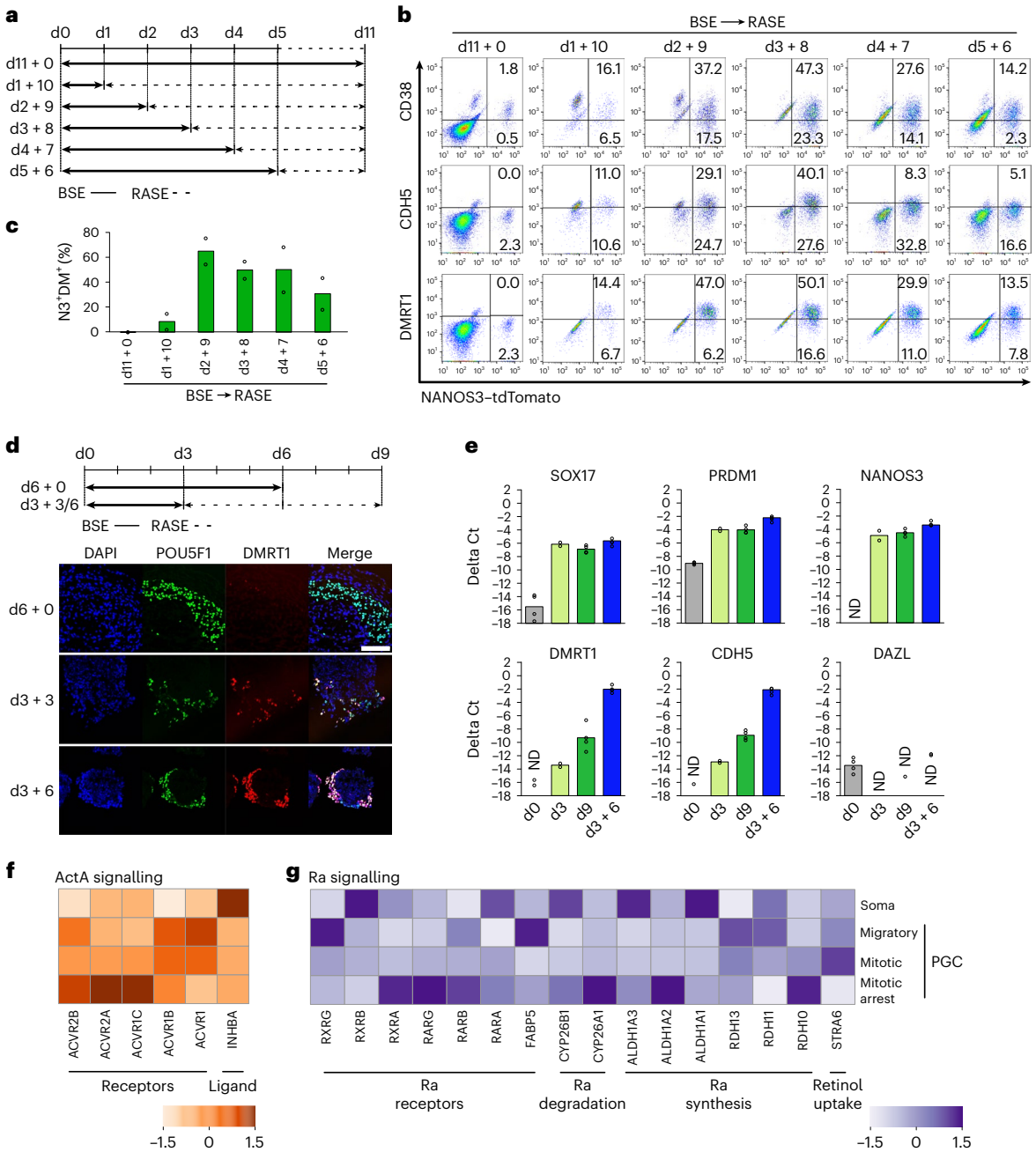


Fig. 2 | Switching signalling in nascent PGCLCs for further development. **a**,**b**, Flow cytometry of NANOS3–tdTomato against CD38–PerCP–Cy5.5, CDH5–AF647 and DMRT1–mVenus. PGCLCs induced from 4i ES cells for 11 days in the presence of BMP2, SCF and EGF (BSE) or BSE replaced with Ra, ActA, SCF and EGF (RASE) on day (d) 1, 2, 3, 4 or 5 (**a**). Percentage population values of NANOS3-positive and CD38, CDH5 or DMRT1-positive and negative cells in the live cell population (**b**). **c**, Quantification for the percentage of NANOS3 and DMRT1 double-positive cells (%N3$^+$DM$^+$). Average from two independent experiments. **d**, Immunofluorescence for POU5F1 and DMRT1 co-staining of aggregates induced from 4i ES cells with 6 days of BSE (day 6), or day 3 (day 3 + 3) or 6 (day 3 + 6) replaced with RASE following 3 days of BSE. Scale bar, 100 µm. **e**, RT–qPCR for TNAP-positive 4i ES cells (d0) and day 3 PGCLCs (d3), and for NANOS3–tdTomato reporter-positive cells induced with BSE for 9 days (d9) and 3 days of BSE followed by RASE for 6 days (d3 + 6) from 4i ES cells. Biologically independent experiments, $n = 4$ for d0, d9, d3 + 6 and $n = 2$ for d3, are shown. Delta Ct values are calculated with housekeeping gene *GAPDH*. ND, some of the value(s) not detected. **f**,**g**, Expression of signalling component of ActA (**f**) and Ra (**g**) from in vivo PGC and soma as transcriptome dataset[24]. The colour codes represent *Z*-score.

A DNA-hypomethylating agent, 5-aza-2′-deoxycytidine (Decitabine, Aza), activated the DAZL reporter but not NANOS3 and DMRT1, confirming DNA demethylation-sensitive reporter expression[39–42] (Extended Data Fig. 3a).

Since PGCs in vivo display increasing levels of DMRT1 before DAZL expression (Fig. 1a), we found that the induction of DMRT1 by dox in ES cells over 8–12 days activated DAZL reporter (dox, day 8: 22.3% and day 12: 34.5%) (Extended Data Fig. 3b,c); replacing BSE with RASE enhanced DAZL (dox + Cy, day 8: 42.6% and day 12: 49.8%) (Extended Data Fig. 3c). Co-induction of DMRT1 with SOX17 by DEX showed more efficient activation of DAZL, but SOX17 alone had no effect (Fig. 3a,b,d and Extended Data Fig. 3d–g). DMRT1/SOX17 repressed

https://doi.org/10.1038/s41556-023-01224-7

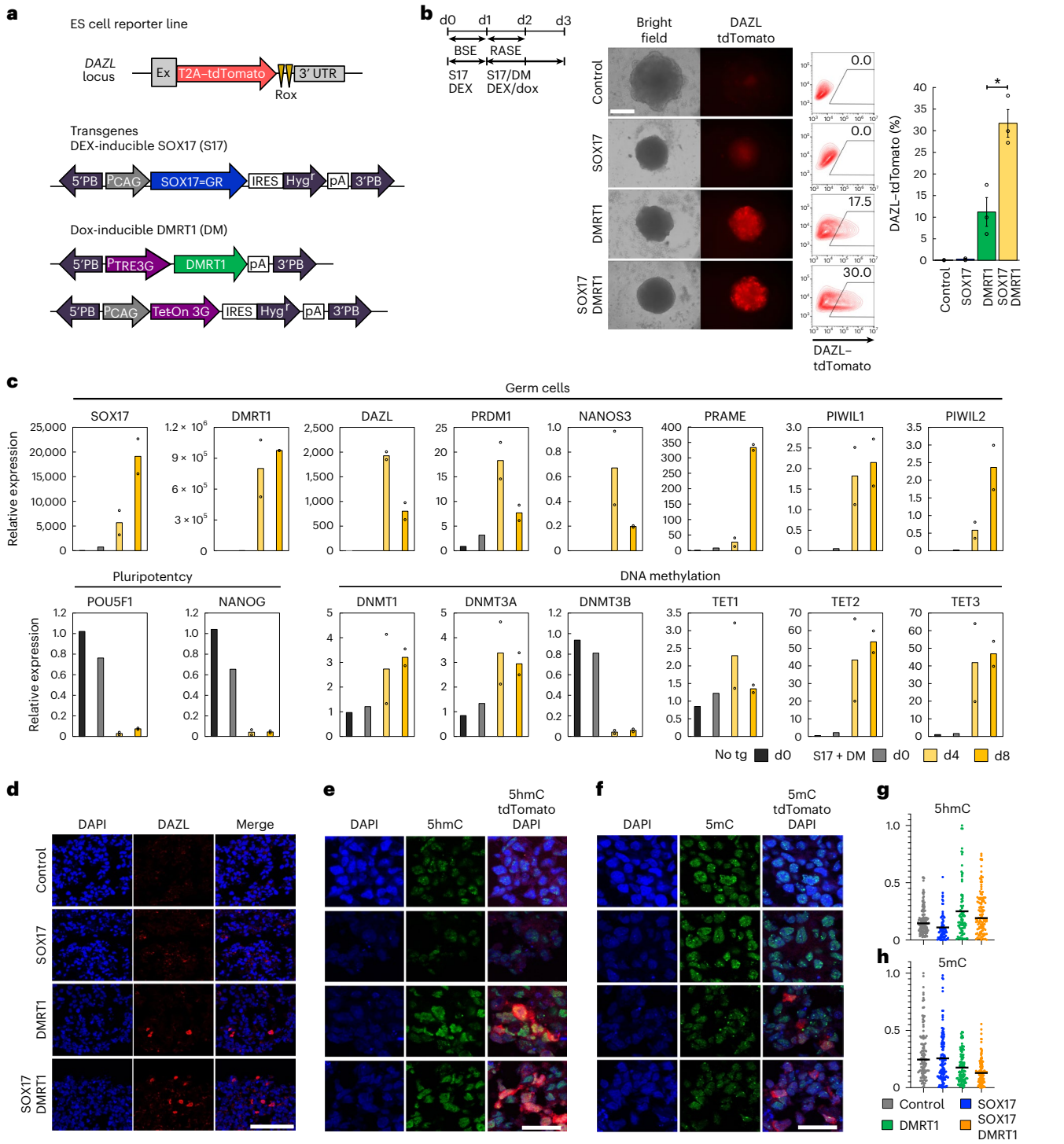


Fig. 3 | DMRT1 activates DAZL expression. a, Expression of DEX-inducible SOX17=GR (S17) and dox-inducible DMRT1 (DM) in DAZL–tdTomato reporter ES cells. **b**, Fluorescence microscope images and flow cytometry of DAZL–tdTomato reporter transgenic male ES cells (WIS2) clone 1 cultured for 3 days with induction of SOX17, DMRT1 or SOX17/DMRT1 in the presence of cytokines: BSE: BMP2/SCF/EGF followed by RASE: Ra/ActA/SCF/EGF. No transgene induction was used as a control. Scale bar, 200 µm. *P = 0.01 calculated using two-tailed t-test. Error bars are mean ± standard error of the mean. Biologically independent experiments, n = 2 for control, SOX17 and n = 3 for DMRT1, SOX17/DMRT1, are shown. **c**, RT–qPCR analysis of DAZL-tdTomato-positive cells induced for 4 and 8 days with the induction of SOX17/DMRT1 transgenes in the presence of cytokines BSE followed by RASE. Values are normalized with housekeeping gene *RPLPO*, and relative values against control samples are presented as mean ± standard deviation. Biologically independent experiments, n = 2 for day (d)4 and d8 and the controls (no tg d0, S17 + DM d0) collected independently, are shown. **d**–**f**, Immunofluorescence for DAZL (**d**), co-staining of 5hmC and tdTomato (**e**) and co-staining of 5mC and tdTomato (**f**) of aggregates induced from transgenic ES cell line with 4 day induction of SOX17, DMRT1 or SOX17/DMRT1. Transgenes were not induced for negative controls. Scale bars, 100 µm (**d**) and 30 µm (**e** and **f**). **g**,**h**, Quantification of immunofluorescence for 5hmC-positive (**g**) and 5mC-positive (**h**) cells of day 4 induction of SOX17, DMRT1 or SOX17/DMRT1 transgenes in DAZL–tdTomato ES cell line (Methods). No transgene induction cells were used as controls. Independent cells analysed for 5hmC: control n = 140, SOX17 n = 80, DMRT1 n = 86, SOX17/DMRT1 n = 120, and 5mC: control n = 113, SOX17 n = 130, DMRT1 n = 109, SOX17/DMRT1 n = 114. The y axis indicates value of scale normalization. The median value is indicated by a black bar in the dot plot. The experiment was repeated independently two times with similar results.

DNA-methyltransferase *DNMT3B* and induced *TET2* and *TET3*, which is analogous to week 7–9 gonadal PGCs (Fig. 3c)[6]. A female ES cell DAZL reporter line with DMRT1/SOX17 responded similarly (Extended Data Figs. 1b,c and 3d,h,i). DAZL-positive cells showed expression of both early PGC (*PRDM1* and *NANOS3*) and gonadal PGC genes (*PRAME*, *PIWIL1* and *PIWIL2*; Fig. 3c and Extended Data Fig. 3i). Notably, *DMRT1*/*SOX17* suppressed pluripotency genes, *POU5F1* and *NANOG*, as seen in male mitotic arrest (week 9) and female pre-meiotic (week 11) PGCs[6,24].

We observed a notable increase in 5hmC upon DMRT1, and DMRT1/SOX17 induction, with a reduction in 5mC, but SOX17 alone had no detectable effect (Fig. 3e–h). Altogether DMRT1 alone and with SOX17 is implicated in the epigenetic programming of the human PGCs, and together induced DAZL-positive PGCLCs (hereafter called DZ$^+$PGCLCs; Fig. 8b). DZ$^+$PGCLCs, when combined with mouse embryonic gonadal cells[28,29], showed DDX4 expression and colonization of testicular tubules, albeit inconsistently (Extended Data Fig. 3j), which, in principle, indicates their developmental potential.

Molecular networks towards human germline commitment

Next, we performed bulk and scRNA-seq for 4i ES cells, NANOS3 (N3)$^+$PGCLC (day 3), DM$^+$PGCLCs and DZ$^+$PGCLCs (day 4 and/or day 8) (Fig. 4). Differentially expressed gene (DEG) analysis of bulk RNA sequencing (RNA-seq) revealed 301 upregulated genes in N3$^+$PGCLC, DM$^+$PGCLCs and DZ$^+$PGCLCs, including 'PGC genes' *SOX17*, *PRDM1*, *TFCP2L1* and *NLRP9* (Methods), 'cellular developmental process' and 'cell fate commitment' by Gene Ontology (GO) (Extended Data Fig. 4a). Upregulated PGC genes include *BRDT*, *BEND4*, *KLF8* and *HDAC4* for N3$^+$PGCLCs, while DM$^+$PGCLCs showed expression of *TCL1A* and *SUSD2*, markers for migratory and mitotic PGCs, respectively (Methods)[24,28], and GO terms 'cell activation' and 'cell migration' (Extended Data Fig. 4a). Mitotic arrest PGC markers *PIWIL1* and *PIWIL2*, as well as GO terms 'sexual reproduction' and 'gamete generation', were upregulated explicitly in DZ$^+$PGCLCs. The commonly downregulated genes related to neurogenesis, cell adhesion and ion transport were identified. For N3$^+$PGCLCs, we found glycan degradation and metabolic pathways, and for DM$^+$PGCLCs we observed sodium ion transport and actin filament-based process. For DZ$^+$PGCLCs, we found cell cycle and chromosome segregation among significantly downregulated DEGs (Extended Data Fig. 4b).

With scRNA-seq analysis, we generated Uniform Manifold Approximation and Projection (UMAP) with Harmony integration (Methods), revealing four clusters (Fig. 4a); *SOX17*, *DMRT1*, *DAZL*, *POU5F1*, *DNMT3B*, *TFCP2L1* and *SUSD2* were enriched in specific clusters in the UMAPs (Fig. 4a and Extended Data Fig. 4c). Label transfer from in vivo male PGCs migratory/mitotic/mitotic arrest[24] to PGCLCs in the UMAP (Fig. 4b) showed 4i ES cells and N3$^+$PGCLCs were labelled mostly with migratory PGCs, while DM$^+$PGCLCs exhibited migratory PGC labels for 5776 cells (similarity score 0.70/1.0) and a subpopulation of mitotic PGC labels for 281 cells (similarity score 0.6/1.0) (Fig. 4b). By contrast, for DZ$^+$PGCLCs, 924 cells were labelled with migratory, 629 cells with mitotic and 1,020 cells with mitotic arrest; the latter two had a higher similarity score >0.55 compared with the score for migratory of 0.31 (Fig. 4b). DZ$^+$PGCLCs also indicated mitotic arrest features due to lacking G2/M and S phases (Methods) (Fig. 4c). Specific marker genes for each in vitro induced group, N3$^+$, DM$^+$ and DZ$^+$ PGCLCs, were reflected in the expression pattern for in vivo PGC developmental phases (Fig. 4d). Consistently, gene set enrichment analysis (GSEA) using bulk RNA-seq revealed significant upregulation of migratory/mitotic PGC genes in DM$^+$PGCLCs, and significant enrichment of mitotic arrest genes in DZ$^+$PGCLCs (Extended Data Fig. 4d). While some mitotic arrest PGC markers, such as *DDX4* and *PIWIL4*, were undetectable (Extended Data Fig. 4e), 24% of mitotic arrest PGC markers were upregulated DEGs in DZ$^+$PGCLCs (Fig. 4e). Interestingly, 33% of the migratory and 39% of the mitotic PGC markers were downregulated in DZ$^+$PGCLCs (Fig. 4e). A heat map provides an overall view of the key transcriptional differences between the three key stages of transitions from nascent to advanced PGCLCs (Fig. 8a). Overall, the PGCLC progression reflects PGC development in vivo towards germline commitment.

DMRT1 targets for human germline commitment

For mechanistic insights concerning DMRT1 and genomic targets, we performed CUT&RUN (C&R)[43–45] for DZ$^+$PGCLCs (on days 4 and 8), which confirmed DMRT1 binding specificity (Extended Data Fig. 5a). Of the 1,148/4,171 protein-coding genes, PGC genes in day 4 DZ$^+$PGCLC were closest to the DMRT1 peaks (henceforth DMRT1 targets). The 630 commonly expressed genes in week 7 and week 9 PGCs include *SOX17*, *PRDM1* and *KLF4*; pluripotency genes, *PRDM14*, *TFCP2L1* and *SUSD2*; epigenetic regulators, *TET1*, *KDM7A* and *KDM4C*; and gonadal PGC genes, *DAZL*, *MAEL*, *PIWIL1* and *PIWIL4* (Fig. 5a). Of these, 149 PGC genes, including *NANOGP8* and *KLF9*, were specific to week 7, while 369 genes, such as *STK33*, *QSER1* and *DDX59*, were specific to week 9 PGCs (Fig. 5a). Over 60% of the DMRT1 peaks for PGC genes were at introns, 32% were intergenic and 1.3% were at promoter transcriptional start sites (TSSs) (Fig. 5b). Intronic and intergenic binding by DMRT1 (47.4% and 43.2%, respectively) was also observed in human testis (Extended Data Fig. 5b)[46].

Combining the C&R with bulk RNA-seq for day 8 DZ$^+$PGCLCs revealed that 13% of DMRT1 target genes were upregulated with GO terms 'Cell fate commitment' and 'sex differentiation', while 27% of the target genes were downregulated with GO terms 'Cell morphogenesis' and 'Brain development' (Extended Data Fig. 5c). We found more DMRT1 targets for mitotic arrest PGC genes when compared with that for migratory and mitotic PGC genes in DZ$^+$PGCLCs (Fig. 5c and Extended Data Fig. 5d). More downregulated target genes were for migratory and mitotic PGCs, than for mitotic arrest PGCs. In contrast, more DMRT1 targets for mitotic arrest genes were upregulated compared with migratory and mitotic genes in DZ$^+$PGCLCs (Fig. 5d and Extended Data Fig. 5e). DMRT1 targets, *CDKN2A* and *CDKN2B* cell cycle inhibitors, were upregulated in DZ$^+$PGCLCs and mitotic arrest PGCs in vivo (Extended Data Fig. 5f,g). Significant enrichment of NR4A1, SMAD2 and ZNF652 binding motifs was found at DMRT1 peaks for migratory/mitotic PGCs genes, while significant enrichment of PGR (progesterone receptor), CEBP:AP1, IRF4 and MYBL2 motifs was enriched for mitotic arrest PGC genes (Fig. 5e), suggesting distinct regulation by DMRT1 for gene suppression in migratory/mitotic PGCs, and gene induction in mitotic arrest PGCs, potentially with a stage-specific binding partner(s).

DNA modification dynamics involving DMRT1

Since the upregulation of 5hmC and downregulation of 5mC was detected by immunofluorescence in DZ$^+$PGCLCs (Fig. 3e–h) as in PGCs in vivo[6], we performed chemical-assisted pyridine borane sequencing plus (CAPS+) for 5hmC, and TET-assisted pyridine borane sequencing with β-glucosyltransferase blocking (TAPSβ) for 5mC to generate single-base-resolution profiling[47,48] in DZ$^+$PGCLCs, 4i ES cells and N3$^+$PGCLCs (Fig. 6).

The global 5hmC levels of 10.1% in DZ$^+$PGCLCs compared with 5.5% for 4i ES cells and 4.2% for N3$^+$PGCLCs (Fig. 6a), while 5mC levels in DZ$^+$PGCLCs, 4i ES cells and N3$^+$PGCLCs were largely similar at ~76–79% (Fig. 6a). Strikingly, however, we identified notable overlapping hyper-5hmC and hypo-5mC differentially methylated regions (DMRs) in DZ$^+$PGCLC (Fig. 6b and Extended Data Fig. 6a), which indicates a dynamic enrichment of 5hmC accompanied by a locus-specific loss of 5mC during differentiation. The hyper-5hmC and hypo-5mC mostly occurred at the DMRT1-bound regions in introns and the intergenic regions (Fig. 5b and Extended Data Figs. 5b and 6b). We observed further enrichment of 5hmC and loss of 5mC at the gene body of mitotic arrest genes compared with migratory and mitotic PGC genes, but not in TSS regions (Fig. 6c). Notably, the transcription factor (TF) binding motifs at the DMRs identified DMRT1 for both hyper-5hmC and hypo-5mC (Fig. 6d). SOX17 was also one of the motifs significantly enriched for both DMRs (Fig. 6d). Consistently, further enrichment of 5hmC and marked depletion of 5mC

https://doi.org/10.1038/s41556-023-01224-7

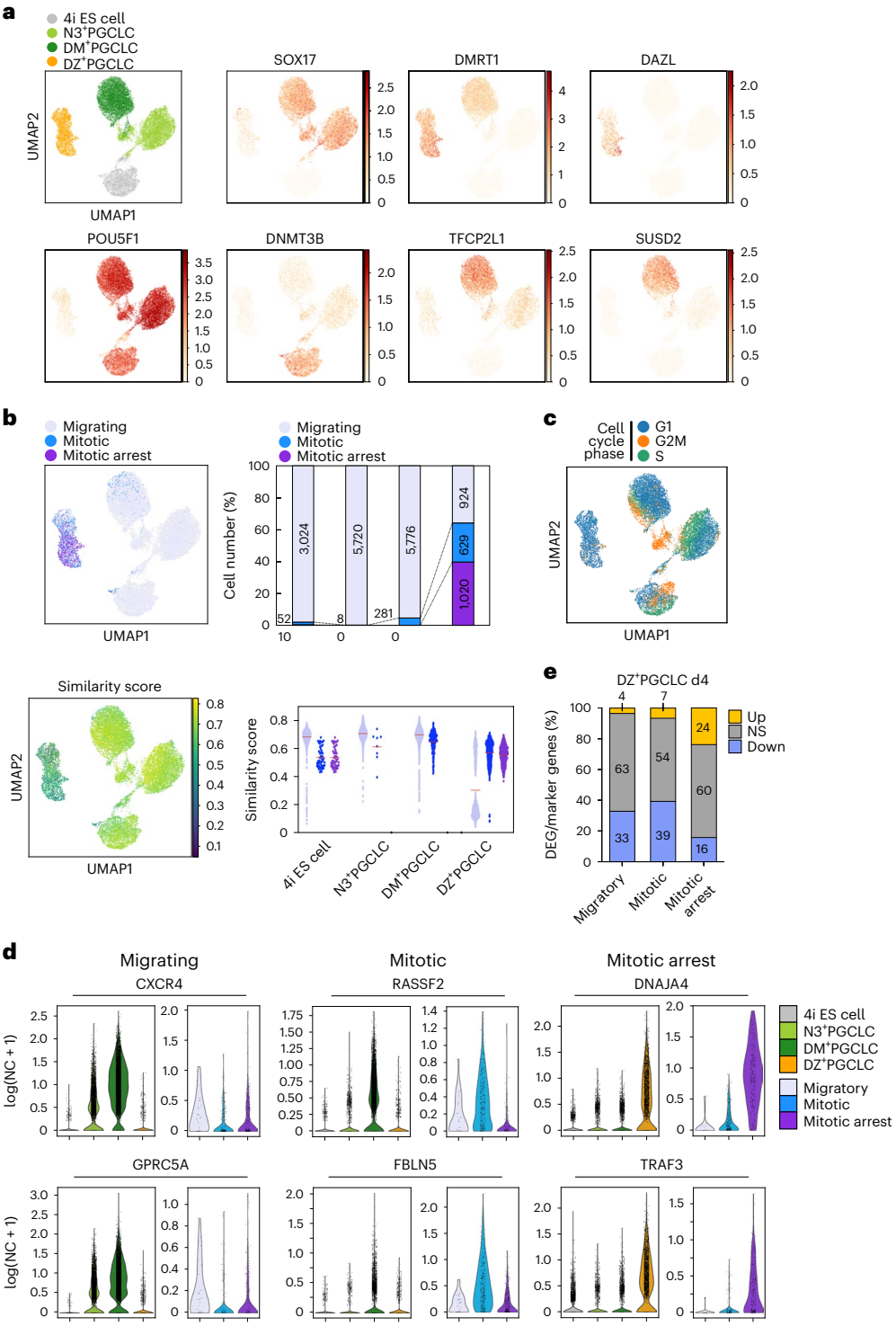


Fig. 4 | Transcriptome networks for human germline commitment in vitro. **a**, UMAPs of 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC day (d)8 scRNA-seq dataset (each from single biological sampling points) batch-corrected with Harmony. Expression of marker genes shown log transformed (normalized count (NC) + 1). SOX17 expression was detected for the endogenous transcripts and not for the chimaeric SOX17=GR transgenes. **b**, Discrete cell type annotations (top) and continuous similarity scores (bottom) transferred with scmap from migratory, mitotic and mitotic arrest PGCs in vivo[24] onto 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC d8 scRNA-seq dataset. UMAPs of 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC d8 scRNA-seq dataset batch-corrected with Harmony and labelled by discrete cell type annotations (top left) and continuous similarity scores (bottom left) from migratory, mitotic and mitotic arrest PGCs in vivo[24]. Top right: bar plot of the proportion and exact number of cells in 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC d8 scRNA-seq dataset predicted to correspond to either migratory, mitotic or mitotic arrest PGCs in vivo. Bottom right: dot plot for the similarity scores between migratory, mitotic or mitotic arrest PGCs in vivo and 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC d8 cells. The mean value is indicated by a red bar in the dot plot. **c**, UMAP of 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC d8 scRNA-seq dataset batch-corrected with Harmony and labelled by phase of the cell cycle. **d**, Marker gene expression for in vitro PGCLCs and migratory, mitotic and mitotic arrest PGCs in vivo. NC, normalized count. **e**, Proportion (*y* axis, %) of DEGs from bulk RNA-seq for day 4 DZ$^+$PGCLCs against 4i ES cells within marker genes for migratory, mitotic or mitotic arrest PGCs in males. The colour codes represent upregulated (Up, yellow), not significant (NS, grey) and downregulated (Down, blue) DEGs. Number of genes in the categories is indicated in bar graphs.

https://doi.org/10.1038/s41556-023-01224-7

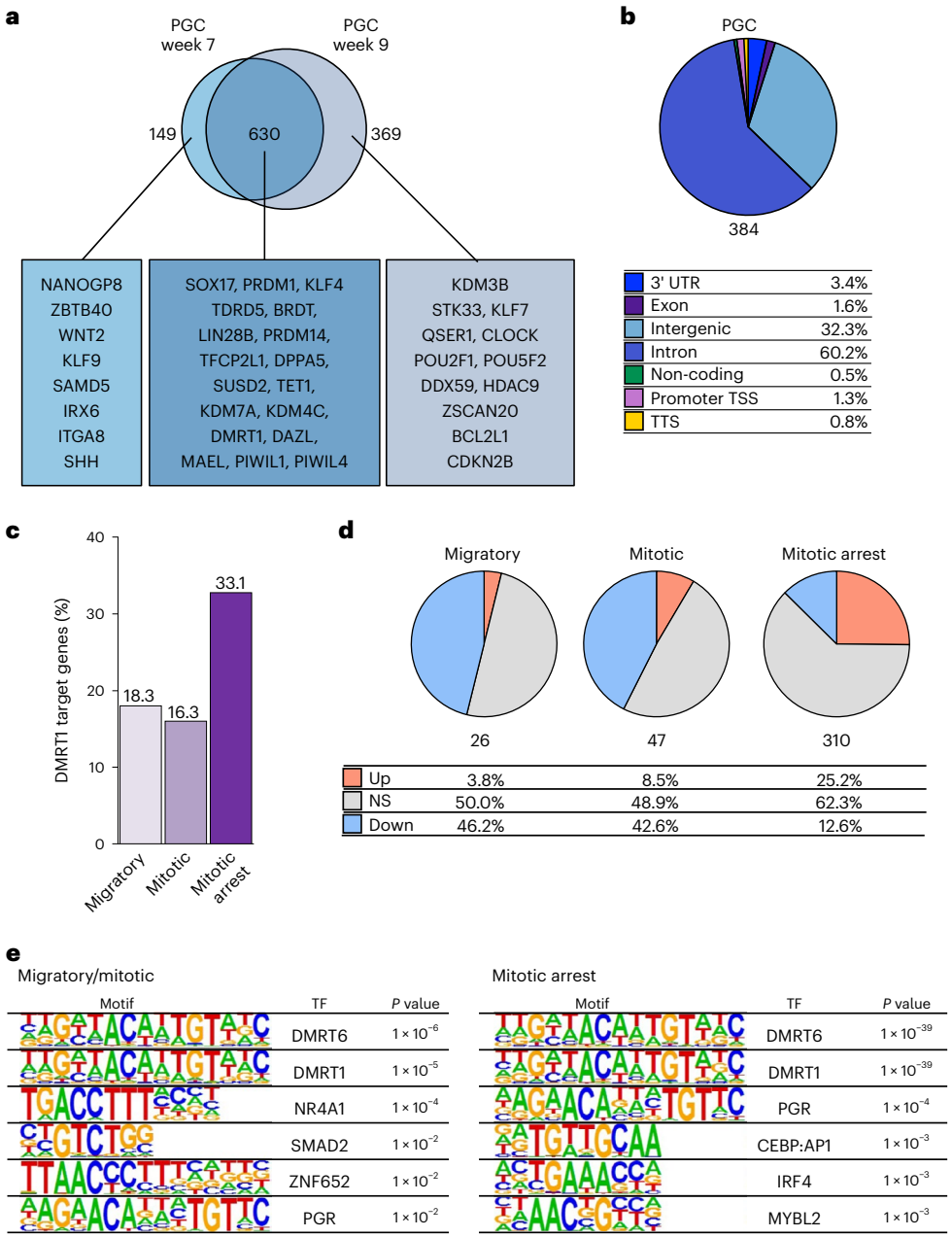


Fig. 5 | DMRT1 targets for foetal germline development. a, Venn diagram for comparing week 7 and week 9 upregulated PGC DEGs that are targets of DMRT1 identified by C&R. Text boxes indicate representative DMRT1 target PGC genes shared or unique in week 7 and week 9 male PGCs. **b**, Pie chart presents the distribution of genomic features for DMRT peaks ($n = 384$) associated with PGC genes. The percentage (%) of the target peaks is shown for each category. **c**, Proportion (*y* axis, %) of DMRT1 targets in d4 DZ$^+$PGCLCs within the marker genes for migratory (18%; 26/142 genes), mitotic (16%: 47/288) or mitotic arrest (33%: 310/937 genes) PGCs. **d**, Proportion of DMRT1 targets for upregulated (Up, orange), not significant (NS, grey) and downregulated (Down, blue) genes for migratory, mitotic or mitotic arrest PGC in d4 DZ$^+$PGCLCs versus 4i ES cells. **e**, Topmost commonly known motifs enriched in DMRT1 peaks. Left: motifs enriched in DMRT1 peaks associated with marker genes for migratory and mitotic male PGCs. Right: motifs enriched in DMRT1 peaks associated with marker genes for mitotic arrest male PGCs. Enrichment *P* values are calculated using binomial distributions. The data in **a**–**e** represent an integration of three biological replicates.

in DZ$^+$PGCLCs was found at DMRT1 binding regions identified by C&R, compared with random regions (Fig. 6e and Extended Data Fig. 6c). A substantial proportion of DMRT1 targets overlapped with hyper-5hmC, hypo-5mC or both (Fig. 6f), suggesting a correlation between DMRT1 genomic binding and DNA methylation changes. Integrative analysis for the methylome, C&R and bulk RNA-seq revealed notable upregulation of DMRT1 targeted mitotic arrest PGC genes with hyper-5hmC in DZ$^+$PGCLCs (Fig. 6g). Conversely, downregulation of DMRT1 targeted migratory and mitotic PGC genes occurred independently of the hyper-5hmC in DZ$^+$PGCLCs (Extended Data Fig. 6d), implying mitotic arrest stage-specific regulation of gene expression by DMRT1 and 5hmC. Genomic loci for upregulated genes in DZ$^+$PGCLCs, such as *DAZL*, *PIWIL2* and *DNAJA4*, displayed multiple sites with hyper-5hmC/hypo-5mC, which co-localized with DMRT1 peaks (Fig. 6h and Extended Data Fig. 6e). Luciferase assay for those regions at the *DAZL* locus with DMRT1 binding motifs showed transcriptional activation in response to DMRT1 (Fig. 6i). Accordingly, DMRT1 binding and consequent DNA methylation changes can

https://doi.org/10.1038/s41556-023-01224-7

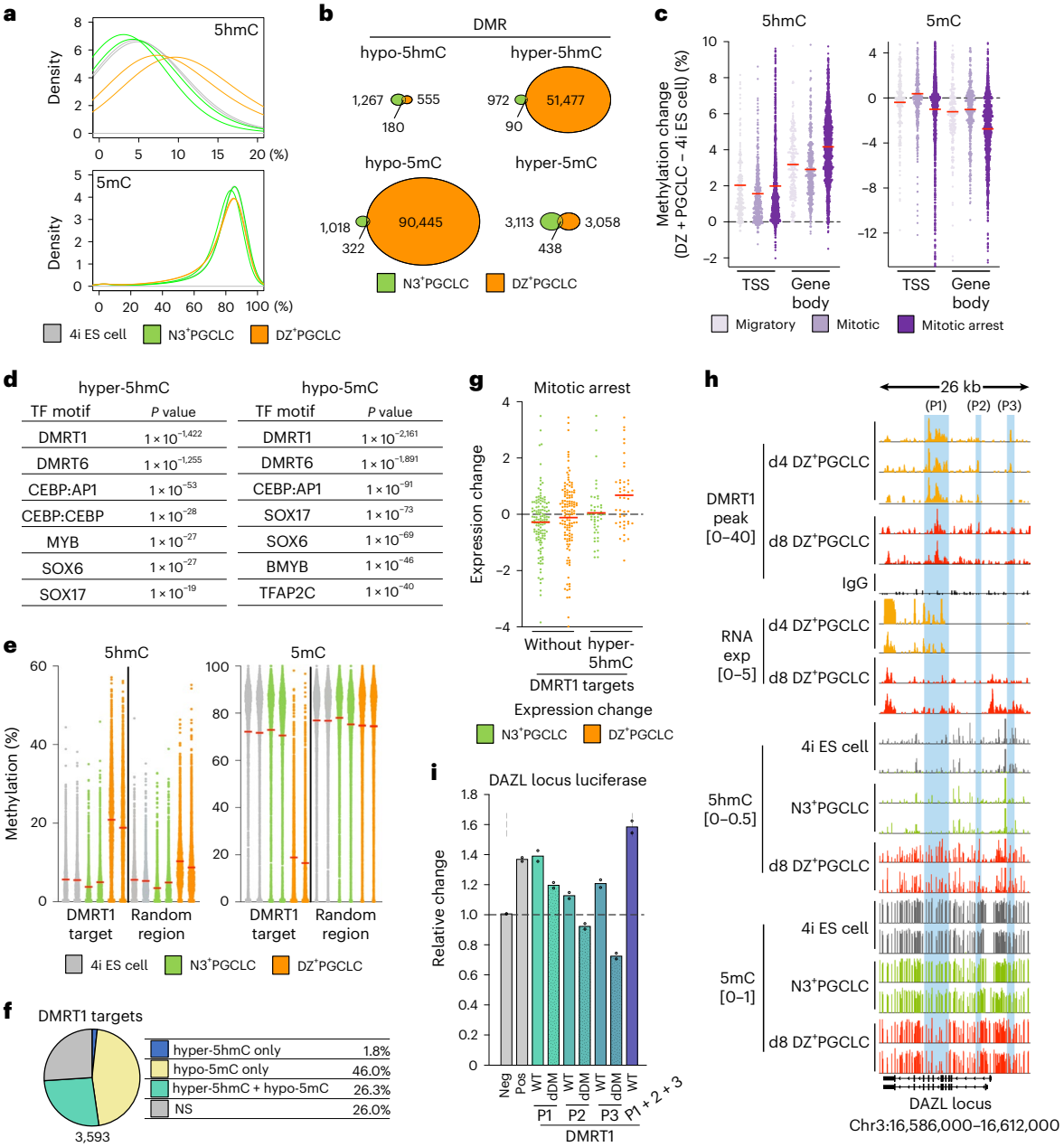


Fig. 6 | DNA methylation profiling and DMRT1 in early human germline. **a**, Percentage of 1 kb tiles (≥3 CpGs, covered >5 times) methylation values across 4i ES cells, N3$^+$PGCLC and day (d)8 DZ$^+$PGCLC methylome. The density plots (bandwidth 0.05) show the distribution of 5hmC or 5mC level across the genome (5hmC, n = 2,657,811 tiles; 5mC, n = 2,622,626 tiles). Two biological replicates of each cell type are shown. **b**, Venn diagrams for DMRs in d8 DZ$^+$PGCLC/4i ES cells and N3$^+$PGCLC/4i ES cells. The data represent an integration of two biological replicates. **c**, DNA methylation of d8 DZ$^+$PGCLC/4i ES cells at approximately −2 to 0 kb TSS and gene body regions for 142 migratory, 288 mitotic and 937 mitotic arrest PGC genes. The mean value is indicated by a red bar in the dot plot. The data represent an integration of two biological replicates. Dashed horizontal line indicates no change in % methylation. **d**, Topmost TF motifs enriched in DMRs for hyper-5hmC (51,567 DMRs) and hypo-5mC (90,767 DMRs) in d8 DZ$^+$PGCLCs/4i ES cells. P values by binomial distributions. **e**, DNA methylation levels of DMRT1 peaks (average length 369 bp, ≥3 CpGs, covered >5 times in all samples) and random regions (n = 3,351). The mean value is indicated by a red bar in the dot plot. Two biological replicates of each cell type are shown. **f**, Percentage of overlap between DMRs and DMRT1 peaks (≥3 CpGs, covered >5 times in all samples). The data represent an integration of two biological replicates. **g**, Expression changes for d8 DZ$^+$PGCLCs/4i ES cells of mitotic arrest PGC genes with DMRT1 peaks with (n = 46 genes) or without (n = 131 genes) hyper-5hmC DMRs in their intronic regions. The mean value is indicated by a line in the dot plot. Dashed horizontal line marks no change in expression. **h**, Visualization for DMRT1 C&R peaks, bulk RNA-seq, 5hmC and 5mC at *DAZL* genomic locus. **i**, Luciferase assay for DMRT1 binding sites at *DAZL*. Peak 1 (P1), 2 (P2) and 3 (P3) indicated in **h** and the deletion of the DMRT1 motif were examined with DMRT1 induction. Empty luciferase vector (Neg: negative control) and that with SV40 enhancer (Pos: positive control). Relative values against Neg. Biologically independent experiments, n = 2. Dashed horizontal line indicates the empty luciferase vector signal activity (Neg sample), set to 1.

regulate stage-specific gene activation. The regulatory mechanism could also involve histone modifications such as the acquisition of H3K36me3 and loss of H3K27me3 with the 5hmC enrichment in gene bodies[49–51], which merits further investigation.

DMRT1 and genomic repeat elements in germline

Expression of genomic repeat elements reflects distinct cell-type-specific features in human germline development[28]. We found downregulation of LTR7 and HERVH-int in DM$^+$PGCLCs and DZ$^+$PGCLCs,

compared with 4i ES cells (Extended Data Fig. 7a–c). There was upregulation of LTR5_Hs and HERVK-int in DM$^+$PGCLCs as in week 5 PGCs (Fig. 7a and Extended Data Fig. 7a,c), whereas SVA_D, LTR12C, ALR/Alpha (a centromeric satellite) and SST1 displayed higher expression in DZ$^+$PGCLCs and week 9 PGCs (Fig. 7a and Extended Data Figs. 7a,c), where mitotic arrest PGCs start to be detectable[24].

We found notable expression of the hominoid-specific LTR12C elements[52] (Fig. 7a), together with the neighbouring genes in day 8 DZ$^+$PGCLCs (Fig. 7b), as in mitotic arrest PGCs in vivo (Fig. 7c). However, genes proximate to SVA_D (another evolutionarily young transposon) did not show transcriptional activation[53] (Fig. 7b,c). There are 75 DMRT1 binding sites at LTR12C in DZ$^+$PGCLCs based on C&R analysis (Fig. 7d,e), and LTR12C bound by DMRT1 had higher 5hmC and lower 5mC consistent with LTR12C expression, compared with those without DMRT1 binding (Fig. 7f and Extended Data Fig. 7d). Accordingly, we reveal a hominoid-specific role of DMRT1 as well as a regulatory role of LTR12C. We also detected 95 DMRT1 peaks at ALR/Alpha centromeric repeats, which showed higher expression in DZ$^+$PGCLC (Fig. 7d,g–i). Luciferase assays confirmed transcriptional activation of ALR/Alpha in response to DMRT1 (Fig. 7j). Together with decreased expression of *CENPA*, a component of the centromere complex[54] in mitotic arrest PGCs and DZ$^+$PGCLCs (Extended Data Fig. 7e), DMRT1 is potentially involved in regulating centromere to control the cell cycle in the human male germline. Our study provides a robust model for mechanistic studies on the role of DMRT1 concerning non-coding genomic regions.

Discussion

We reveal a critical role of DMRT1 in epigenetic resetting and transcription regulation during human germ cell lineage commitment. Specification of PGCs (approximately weeks 2–3) is followed by a multi-step process towards irreversible commitment and gain of competence for gametogenesis[25,42,55,56]. The functional role of DMRT1 in vivo and in vitro coincides with the gain of 5hmC and loss of 5mC, which is exemplified by the expression of DAZL, a DNA methylation-sensitive gene (Fig. 8b)[6,24]. DMRT1 expression commences in migrating male and female PGCs, with a progressive increase in DAZL, which marks germline commitment; later, expression of DMRT1 in germ cells is restricted to males[27,57–60]. Our in vitro model enables mechanistic studies that are hampered by limited access to week 2–6 human embryos.

A transient BMP signal is essential and sufficient for PGCLC specification following the induction of SOX17 and PRDM1 (refs. 11,14,61). However, the sustained presence of BMP interferes with DMRT1 expression in PGCLCs; replacing it with ActA and Ra after specification induces DMRT1 and CDH5, respectively, marking the progression of PGCLCs to the migratory stage. The post-specification signalling system comprising ActA and Ra is present in somatic tissues along the migratory route surrounding the PGCs in vivo (Fig. 2f,g). ActA is also crucial for differentiating mesendoderm, yolk sac endoderm and definitive endoderm forming hindgut later[62–64], and Ra regulates anterior–posterior patterning with the gradient expression[65,66]. Activin and Ra have a role in migrating PGCs in mice[67–70]. CDH5, a homophilic adhesion molecule[71], may promote interactions between PGCs and surrounding tissues; notably, CDH5 is not detectable in mouse or cynomolgus monkey PGCs[17].

DMRT1 induces the expression of DNA methylation-sensitive PGC genes, *PIWIL2* and *DAZL*[39–41] through converting 5mC to 5hmC in DZ$^+$PGCLCs, which is evident in PGCs in vivo[6]. DMRT1 also causes the downregulation of *DNMT3B* (but not *DNMT3A*) and upregulation of *TET2* (but not *TET1*). In mice, DNMT3B is required for DNA methylation at gene bodies, while DNMT3A is preferentially recruited to enhancers/promoters[72–76]. Tet2 depletion in mice reduces 5hmC in gene bodies, while Tet1 depletion decreases 5hmC at promoters[77], which is also reflected in the gain of 5hmC/loss of 5mC at gene bodies in DZ$^+$PGCLCs. We observed a localized enhancement of 5hmC and a decrease in 5mC with DMRT1 binding. Similarly, localized DNA demethylation occurs by recruitment of TET1 and TET2 by FOXA1, a pioneer TF, to its binding sites[78]. The direct targets of DMRT1, *TET1* and *QSER1*, protect DNA hypomethylated regions from de novo methylation[79]. Further mechanistic studies are required to elucidate the molecular regulation for the site-specific 5hmC/5mC by DMRT1. SOX17 and PRDM1 also contribute to the epigenetic reprogramming through the upregulation of *TET1*/*TET2*, histone demethylases and the repression of *DNMT3B*[6,61]. Unlike in mice, DAZL expression commences in migrating and proliferating human PGCs[27], which is incompatible with the onset of meiotic gene expression; these aspects may be regulated differently in humans. DMRT1 binding to hominoid-specific LTR12C with the 5hmC/5mC epigenetic changes also reflect species differences; the upregulation of some neighbouring genes, including mitotic arrest genes, suggests their regulatory role in human PGCs. Activation of SVA_D, another evolutionarily young transposable element[53], occurred without detectable upregulation of neighbouring genes. The evolution of the transcriptional regulation by DMRT1 through unique repeat elements for germ cell development merits further investigation.

DMRT1 expression occurs both in PGCs and gonadal somatic cells, but global DNA demethylation occurs only in PGCs[6–9,24]. In this context, SOX17 expression is restricted to germ cells, suggesting the combination of SOX17/DMRT1 contributes to DNA demethylation as we found both motifs at hyper-5hmC/hypo-5mC in DZ$^+$PGCLCs. Tumours induced by the Yamanaka pluripotency factors (OSKM) exhibit DNA hypomethylation and DAZL expression through DMRT1 (ref. 80), suggesting that a combinatory effect of DMRT1 with other factors, such as pluripotency genes, enables regulation of DNA methylation for germline commitment. Dmrt1 is also present in rodent germline[58,59] but without Sox17 (refs. 13,19–21), indicating potential mechanistic and functional species differences[81]. The mitotic arrest of PGC/PGCLCs might also, in part, be induced by DMRT1 following the migratory and mitotic phase of PGCs. DMRT1 dosage probably increases progressively during human germline development, reaching higher levels during the mitotic arrest of PGCs (Fig. 1a)[24]. Regulation of DMRT1 levels might be crucial for the stage-specific role in the human foetal germline. A dose-sensitive regulation by DMRT1 has been reported in another context[59,82].

DMRT1 repressed pluripotency factors, a hallmark of germline commitment towards mitotic arrest in male PGCs and pre-meiotic PGCs in female[24] (Fig. 8b); suppression of DMRT1 in second-trimester human foetal testis induces upregulation of POU5F1 (ref. 83), and Dmrt1 depletion in 129Sv male mice causes germ cell teratoma through the lack of suppression of pluripotency genes[59,84]. Notably, genomic variants were identified near the DMRT1 locus in testicular germ cell tumours, indicating the potential origin of carcinoma in situ, and the role of DMRT1 for irreversible germline commitment[58,85–87].

Our in vitro gene induction system did not allow an accurate temporal and dosage control of DMRT1 as is likely in vivo. Accordingly, DZ$^+$PGCLCs exhibit enrichment of 5hmC with incomplete global DNA demethylation equivalent to the epigenetic status of migratory PGCs; instead, there was induction of the mitotic arrest transcriptional programme. A gradual increase in DMRT1 in vivo occurs over a few weeks leading step-wise towards mitotic arrest PGCs[24]. The high levels of DMRT1 in our system may trigger mitotic arrest more rapidly within 4–8 days, which potentially prevents DNA replication coupled dilution of 5mC in PGCLCs[88,89]. We also did not observe transcriptional repression of H3K9 methyltransferases, G9A and SETDB1 in DZ$^+$PGCLCs, which accompany global DNA demethylation in PGCs in vivo[10,24]. Future studies with exquisitely tunable expression of DMRT1 are warranted to mimic the stage-specific role. Note that the lack DDX4 expression in DZ$^+$PGCLCs as in in vivo mitotic arrest PGCs may occur if combined with gonadal somatic tissues (Extended Data Fig. 3j).

While our mechanistic studies on DMRT1 were performed using male cells, female DZ$^+$PGCLCs also showed downregulation

https://doi.org/10.1038/s41556-023-01224-7

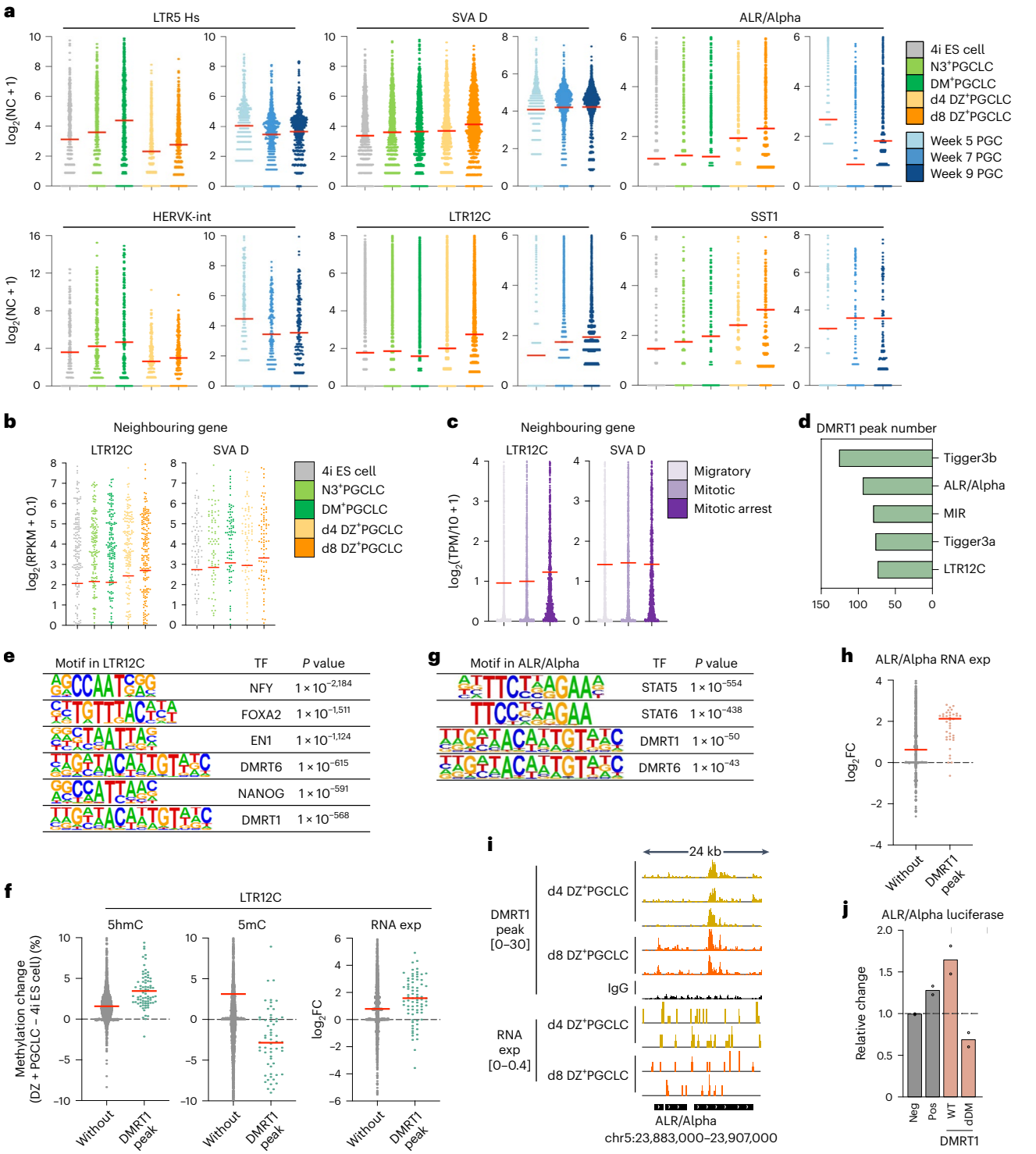


Fig. 7 | Regulation of genomic repeat elements and DMRT1 in early human germline. **a**, Expression of genomic repeats. The mean value is indicated by a red bar in the dot plot. NC, normalized count. **b**, Expression of genes nearest LTR12C (n = 206 genes) and SVA_D (n = 95 genes). The mean value is indicated by a red bar in the dot plot. **c**, Expression of genes nearest LTR12C (n = 1,093 genes) or SVA_D (n = 784 genes) in migratory/mitotic/mitotic arrest male PGCs. The mean value is indicated by a red bar in the dot plot. **d**, Top repeat subfamilies with DMRT1 peaks in day (d)8 DZ$^+$PGCLCs. **e**, Topmost TF motifs enriched in LTR12C (n = 2,816). P values by binomial distributions. **f**, Methylation (5hmC/5mC) and RNA expression (exp) changes of LTR12C with (n = 77) or without (n = 2,739) DMRT1 peaks in d8 DZ$^+$PGCLCs/4i ES cells. The mean value is indicated by a red bar i n the dot plot. Dashed horizontal line indicates no change in % methylation (5hmC and 5mC) or RNA expression levels (RNA exp). FC, fold changes. **g**, Topmost TF motifs at ALR/Alpha (n = 1,912). P values by binomial distributions. **h**, Expression changes of ALR/Alpha with (n = 41) or without (n = 1,871) DMRT1 peaks in d8 DZ$^+$PGCLCs/4i ES cells. The mean value is indicated by a red bar in the dot plot. Dashed horizontal line marks no change in ALR/Alpha RNA expression. FC, fold changes. The data in **a**–**h** represent an integration of two biological replicates. **i**, DMRT1 CR peaks and RNA-seq at ALR/Alpha. **j**, Luciferase assay for ALR/Alpha and that with DMRT1 motif deletion (dDM) after DMRT1 induction. Empty luciferase vector (Neg: negative control) and that with SV40 enhancer (Pos: positive control). Relative values against Neg. Independent experiments, n = 2. Dashed horizontal line indicates the empty luciferase vector signal activity (Neg sample), set to 1.

https://doi.org/10.1038/s41556-023-01224-7

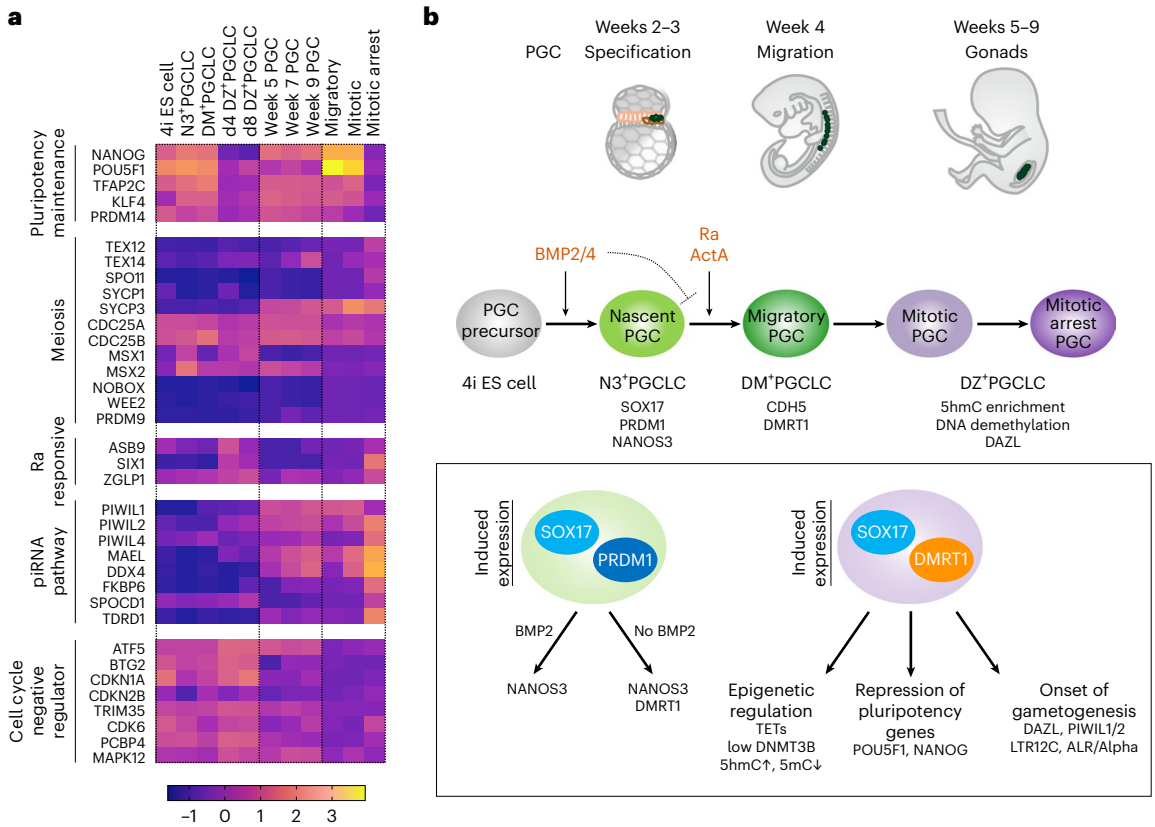


Fig. 8 | Molecular regulation of human germline commitment. a, Heat map presents gene expression *Z*-score of the sample set indicated for enriched genes in 4i ES cells (4i), N3$^+$PGCLCs (N3), DM$^+$PGCLCs (DM), DZ$^+$PGCLCs (day (d)4, 4DZ; d8, 8DZ), week 5/7/9 PGCs, migratory, mitotic and mitotic arrest PGCs. piRNA, PIWI-interacting RNA. The data show an integration of two biological replicates. **b**, Summary scheme for the molecular regulation of human germline commitment. Human PGC specification (approximately weeks 2–3) from the precursors (4i ES cells) in response to BMP2 or BMP4, which induces expression of SOX17 and PRDM1, followed by NANOS3 (N3$^+$PGCLCs). Further development occurs by switching the signalling from BMP to Ra and ActA, marked by the expression of CDH5 and DMRT1 (DM$^+$PGCLCs), characteristics for migratory/mitotic PGCs (approximately weeks 4–5). Migratory PGCs undergo progressive DNA demethylation through transient 5hmC enrichment, followed by the expression of DAZL (DZ$^+$PGCLCs), indicating germline commitment characterized as mitotic arrest status in males (approximately weeks 5–9). Induced expression of SOX17 and PRDM1 can induce *NANOS3*, but the expression of DMRT1 occurs only in the absence of BMP2, suggesting the suppressive effect on DMRT1. Induced expression of DMRT1 and SOX17 modulates the epigenetic programme, including the enrichment of 5hmC and loss of 5mC directly bound by DMRT1, and the potential regulators, *TETs* and *DNMT3B*. DMRT1 is also involved in the suppression of pluripotency genes and induction of later PGC programme, *DAZL*, *PIWIL1* and *PIWIL2*, as well as specific repeat elements, LTR12C and ALR/Alpha, collectively the hallmarks of germline commitment towards the onset of gametogenesis.

of pluripotency genes, *POU5F1* and *NANOG* (Extended Data Fig. 3i), indicating the RA-responsive phase in female. Expression of RA-responsive genes occurs in male DZ$^+$PGCLCs and mitotic arrest PGCs[24]. DMRT1 declines in female PGCs with meiotic entry, following the RA-responsive stage[24]. Accordingly, there is probably a common role for DMRT1 for the germline commitment in males and females; DMRT1 downregulation may lead to meiosis in the female germline, but further studies with respect to female specific events, such as the status of the X chromosome, merit consideration in the future.

Our study implicates DMRT1 as an essential factor regulating the transition from nascent PGCs to gametogenesis-competent cells, involving locus-specific epigenetic resetting. We provide a critical foundation for further investigations and experimental approaches for advances in human germline biology and in vitro gametogenesis.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-023-01224-7.

References

1. Seydoux, G. & Braun, R. E. Pathway to totipotency: lessons from germ cells. *Cell* **127**, 891–904 (2006).
2. Felici, M. D. Origin, migration, and proliferation. *Oogenesis* https://doi.org/10.1007/978-0-85729-826-3_2 (2012).
3. Tang, W. W. C., Kobayashi, T., Irie, N., Dietmann, S. & Surani, M. A. Specification and epigenetic programming of the human germ line. *Nat. Rev. Genet.* **17**, 585–600 (2016).
4. Saitou, M. & Hayashi, K. Mammalian in vitro gametogenesis. *Science* **374**, eaaz6830 (2021).
5. Fujimoto, T., Miyayama, Y. & Fuyuta, M. The origin, migration and fine morphology of human primordial germ cells. *Anat. Rec.* **188**, 315–329 (1977).
6. Tang, W. W. C. et al. A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161**, 1453–1467 (2015).
7. Gkountela, S. et al. DNA demethylation dynamics in the human prenatal germline. *Cell* **161**, 1425–1436 (2015).
8. Guo, F. et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161**, 1437–1452 (2015).

9. Guo, H. et al. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res.* **27**, 165–183 (2017).
10. Gruhn, W. H. et al. Epigenetic resetting in the human germ line entails histone modification remodeling. *Sci. Adv.* **9**, eade1257 (2023).
11. Sasaki, K. et al. Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* **17**, 178–194 (2015).
12. Yao, C., Yao, R., Luo, H. & Shuai, L. Germline specification from pluripotent stem cells. *Stem Cell Res. Ther.* **13**, 74 (2022).
13. Kobayashi, T. et al. Principles of early human development and germ cell program from conserved model systems. *Nature* **546**, 416–420 (2017).
14. Irie, N. et al. SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253–268 (2015).
15. Irie, N., Sybirna, A. & Surani, M. A. What can stem cell models tell us about human germ cell biology? *Curr. Top. Dev. Biol.* **129**, 25–65 (2018).
16. Kobayashi, T. et al. Tracing the emergence of primordial germ cells from bilaminar disc rabbit embryos and pluripotent stem cells. *Cell Rep.* **37**, 109812 (2021).
17. Sasaki, K. et al. The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Dev. Cell* **39**, 169–185 (2016).
18. Seita, Y. et al. Efficient generation of marmoset primordial germ cell-like cells using induced pluripotent stem cells. *eLife* **12**, e82263 (2023).
19. 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**, 427–439 (2009).
20. Kanai-Azuma, M. et al. Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129**, 2367–2379 (2002).
21. Oikawa, M. et al. Functional primordial germ cell–like cells from pluripotent stem cells in rats. *Science* **376**, 176–179 (2022).
22. Kobayashi, T. et al. Germline development in rat revealed by visualization and deletion of Prdm14. *Development* **147**, dev183798 (2020).
23. Garcia-Alonso, L. et al. Single-cell roadmap of human gonadal development. *Nature* **607**, 540–547 (2022).
24. Li, L. et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* **20**, 858–873.e4 (2017).
25. Nicholls, P. K. et al. Mammalian germ cells are determined after PGC colonization of the nascent gonad. *Proc. Natl Acad. Sci. USA* **116**, 25677–25687 (2019).
26. Oosterhuis, J. W. & Looijenga, L. H. J. Human germ cell tumours from a developmental perspective. *Nat. Rev. Cancer* **19**, 522–537 (2019).
27. Alves-Lopes, J. P. et al. Specification of human germ cell fate with enhanced progression capability supported by hindgut organoids. *Cell Rep.* **42**, 111907 (2023).
28. Hwang, Y. S. et al. Reconstitution of prospermatogonial specification in vitro from human induced pluripotent stem cells. *Nat. Commun.* **11**, 5656 (2020).
29. Yamashiro, C. et al. Generation of human oogonia from induced pluripotent stem cells in vitro. *Science* **362**, 356–360 (2018).
30. Canovas, S., Campos, R., Aguilar, E. & Cibelli, J. B. Progress towards human primordial germ cell specification in vitro. *Mol. Hum. Reprod.* **23**, 4–15 (2016).
31. Matson, C. K. & Zarkower, D. Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nat. Rev. Genet.* **13**, 163–174 (2012).
32. Zarkower, D. Current topics in developmental biology—DMRT genes in vertebrate gametogenesis. *Curr. Top. Dev. Biol.* **102**, 327–356 (2013).
33. Gonen, N. & Lovell-Badge, R. The regulation of Sox9 expression in the gonad. *Curr. Top. Dev. Biol.* **134**, 223–252 (2019).
34. Dolci, S. et al. Requirement for mast cell growth factor for primordial germ cell survival in culture. *Nature* **352**, 809–811 (1991).
35. Matsui, Y. et al. Effect of Steel factor and leukaemia inhibitory factor on murine primordial germ cells in culture. *Nature* **353**, 750–752 (1991).
36. Ohinata, Y. et al. A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571–584 (2009).
37. Hohoff, C., Börchers, T., Rüstow, B., Spener, F. & van Tilbeurgh, H. Expression, purification, and crystal structure determination of recombinant human epidermal-type fatty acid binding protein. *Biochemistry* **38**, 12229–12239 (1999).
38. Smathers, R. L. & Petersen, D. R. The human fatty acid-binding protein family: evolutionary divergences and functions. *Hum. Genomics* **5**, 170 (2011).
39. Hackett, J. A. et al. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139**, 3623–3632 (2012).
40. Hargan-Calvopina, J. et al. Stage-specific demethylation in primordial germ cells safeguards against precocious differentiation. *Dev. Cell* **39**, 75–86 (2016).
41. Maatouk, D. M. et al. DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411–3418 (2006).
42. Nicholls, P. K. & Page, D. C. Germ cell determination and the developmental origin of germ cell tumors. *Development* **148**, dev198150 (2021).
43. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *eLife* **8**, e46314 (2019).
44. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
45. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
46. Murphy, M. W. et al. An ancient protein–DNA interaction underlying metazoan sex determination. *Nat. Struct. Mol. Biol.* **22**, 442–451 (2015).
47. Liu, Y. et al. Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. *Nat. Commun.* **12**, 618 (2021).
48. Xu, H. et al. Modular oxidation of cytosine modifications and their application in direct and quantitative sequencing of 5-hydroxymethylcytosine. *J. Am. Chem. Soc.* **145**, 7095–7100 (2023).
49. Greco, C. M. et al. DNA hydroxymethylation controls cardiomyocyte gene expression in development and hypertrophy. *Nat. Commun.* **7**, 12418 (2016).
50. Hahn, M. A. et al. Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Rep.* **3**, 291–300 (2013).
51. Wu, H. et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Gene Dev.* **25**, 679–684 (2011).
52. Trizzino, M. et al. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **27**, 1623–1633 (2017).
53. Wang, H. et al. SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.* **354**, 994–1007 (2005).
54. Amor, D. J., Kalitsis, P., Sumer, H. & Choo, K. H. A. Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol.* **14**, 359–368 (2004).

55. Hu, Y.-C. et al. Licensing of primordial germ cells for gametogenesis depends on genital ridge signaling. *PLoS Genet.* **11**, e1005019 (2015).
56. McLaren, A. Primordial germ cells in the mouse. *Dev. Biol.* **262**, 1–15 (2003).
57. Lei, N. et al. Sex-specific differences in mouse DMRT1 expression are both cell type- and stage-dependent during gonad development. *Biol. Reprod.* **77**, 466–475 (2007).
58. Krentz, A. D. et al. DMRT1 promotes oogenesis by transcriptional activation of Stra8 in the mammalian fetal ovary. *Dev. Biol.* **356**, 63–70 (2011).
59. Krentz, A. D. et al. The DM domain protein DMRT1 is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency. *Proc. Natl Acad. Sci. USA* **106**, 22323–22328 (2009).
60. Jørgensen, A., Nielsen, J. E., Jensen, M. B., Græm, N. & Meyts, E. R.-D. Analysis of meiosis regulators in human gonads: a sexually dimorphic spatio-temporal expression pattern suggests involvement of DMRT1 in meiotic entry. *Mol. Reprod. Med.* **18**, 523–534 (2012).
61. Tang, W. W. C. et al. Sequential enhancer state remodelling defines human germline competence and specification. *Nat. Cell Biol.* **24**, 448–460 (2022).
62. D'Amour, K. A. et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.* **23**, 1534–1541 (2005).
63. Loh, K. M. et al. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell* **14**, 237–252 (2014).
64. Mackinlay, K. M. et al. An in vitro stem cell model of human epiblast and yolk sac interaction. *eLife* **10**, e63930 (2021).
65. Duester, G. Retinoic acid synthesis and signaling during early organogenesis. *Cell* **134**, 921–931 (2008).
66. Sakai, Y. et al. The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. *Gene Dev.* **15**, 213–225 (2001).
67. Lopes, S. M. C. et al. Altered primordial germ cell migration in the absence of transforming growth factor β signaling via ALK5. *Dev. Biol.* **284**, 194–203 (2005).
68. Senft, A. D., Bikoff, E. K., Robertson, E. J. & Costello, I. Genetic dissection of Nodal and Bmp signalling requirements during primordial germ cell development in mouse. *Nat. Commun.* **10**, 1089 (2019).
69. Koshimizu, U., Watanabe, M. & Nakatsuji, N. Retinoic acid is a potent growth activator of mouse primordial germ cells in vitro. *Dev. Biol.* **168**, 683–685 (1995).
70. Ohta, H. et al. In vitro expansion of mouse primordial germ cell-like cells recapitulates an epigenetic blank slate. *EMBO J.* **36**, 1888–1907 (2017).
71. Giannotta, M., Trani, M. & Dejana, E. VE-cadherin and endothelial adherens junctions: active guardians of vascular integrity. *Dev. Cell* **26**, 441–454 (2013).
72. Charlton, J. et al. TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet.* **52**, 819–827 (2020).
73. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
74. Liao, J. et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* **47**, 469–478 (2015).
75. Manzo, M. et al. Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *EMBO J.* **36**, 3421–3434 (2017).
76. Jeong, M. et al. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet* **46**, 17–23 (2014).
77. Huang, Y. et al. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **111**, 1361–1366 (2014).
78. Lemma, R. B. et al. Pioneer transcription factors are associated with the modulation of DNA methylation patterns across cancers. *Epigenet. Chromatin* **15**, 13 (2022).
79. Dixon, G. et al. QSER1 protects DNA methylation valleys from de novo methylation. *Science* **372**, eabd0875 (2021).
80. Taguchi, J. et al. DMRT1-mediated reprogramming drives development of cancer resembling human germ cell tumors with features of totipotency. *Nat. Commun.* **12**, 5041 (2021).
81. Hansen, C. L. & Pelegri, F. Primordial germ cell specification in vertebrate embryos: phylogenetic distribution and conserved molecular features of preformation and induction. *Front. Cell Dev. Biol.* **9**, 730332 (2021).
82. Ioannidis, J. et al. Primary sex determination in birds depends on DMRT1 dosage, but gonadal sex does not determine adult secondary sex characteristics. *Proc. Natl Acad. Sci. USA* **118**, e2020909118 (2021).
83. Macdonald, J. et al. DMRT1 repression using a novel approach to genetic manipulation induces testicular dysgenesis in human fetal gonads. *Hum. Reprod.* **33**, 2107–2121 (2018).
84. Krentz, A. D. et al. Interaction between DMRT1 function and genetic background modulates signaling and pluripotency to control tumor susceptibility in the fetal germ line. *Dev. Biol.* **377**, 67–78 (2013).
85. Looijenga, L. H. J. et al. Genomic and expression profiling of human spermatocytic seminomas: primary spermatocyte as tumorigenic precursor and DMRT1 as candidate chromosome 9 gene. *Cancer Res.* **66**, 290–302 (2006).
86. Poynter, J. N., Hooten, A. J., Frazier, A. L. & Ross, J. A. Associations between variants in KITLG, SPRY4, BAK1, and DMRT1 and pediatric germ cell tumors. *Genes Chromosomes Cancer* **51**, 266–271 (2012).
87. Turnbull, C. et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat. Genet* **42**, 604–607 (2010).
88. Hackett, J. A. et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–452 (2012).
89. 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**, 340–353 (2012).

Methods

Ethics statement

Human embryonic tissues were used with permission from the National Health Service Research Ethical Committee, UK (Research Ethics Committee number 96/085). Patients (who had already decided to undergo the termination of pregnancy operation) fully and freely consented to donate the foetal tissues for medical and academic research. We received genital ridges and dissected to isolate gonads from mesonephric tissues. The gonadal tissues were dissociated into single-cell suspension with Collagenase IV (2.6 mg ml^{-1}) (Sigma, C5138) and DNase I (10 U ml^{-1}) in Dulbecco's modified Eagle medium (DMEM)–F/12 (Gibco). Cells were resuspended in fluorescence-activated cell sorting (FACS) medium (phosphate-buffered saline (PBS) with 3% foetal calf serum) with 5 µl of Alexa Fluor 488 anti-alkaline phosphatase (BD Pharmingen, 561495) and 5 µl of PerCP–Cy5.5 anti-CDH5 (BD Pharmingen, 561566) antibodies for flow cytometry. Medical or surgical termination of pregnancy was carried out at Addenbrooke's Hospital, Cambridge, UK. This study did not involve the use of human gametes, pre-implantation embryos or experimental models mimicking early human development. Where applicable, our study is compliant with the International Society for Stem Cell Research guidelines. All samples were handled and stored according to the Human Tissue Act regulations. The Gurdon Institute safety committee carried out appropriate scrutiny, including risk assessments.

Collection of PGCs from human embryos

Crown–rump length and anatomical features, including limb and digit development, were used to determine the developmental stage of human embryos with reference to Carnegie staging. The sex of embryos was determined by sex determination PCR as previously described[90]. Genital ridges were dissected and separated from surrounding mesonephric tissues and dissociated into single-cell suspension with Collagenase IV (2.6 mg ml^{-1}) (Sigma, C5138) and DNase I (10 U ml^{-1}) in DMEM–F/12 (Gibco) at 37 °C for 15–30 min. Cells were resuspended in FACS medium (PBS with 3% foetal calf serum) with 5 µl of Alexa Fluor 488 anti-alkaline phosphatase (BD Pharmingen, 561495) and 5 µl of PerCP–Cy5.5 anti-CDH5 (BD Pharmingen, 561566) antibodies for 20 min at room temperature. Flow cytometry was performed with BD LSRFortessa Cell Analyzer (BD Biosciences), and dot plots were generated by FlowJo software.

Cell culture

Approval for the use of all ES cell lines used in this study was granted by the MRC Steering Committee for the UK Stem Cell Bank and for the Use of Stem Cell Lines. Male ES cell line, WIS2 (46XY), was kindly provided by Weizmann Institute of Science, Israel[91]. Female ES cell line, Shef-6 (46XX), was obtained from the UK Stem Cell Bank (UKSCB accession no. R-05-031). 4i ES cells were maintained on irradiated mouse embryonic fibroblasts (MEFs) (purchased from MTI-GlobalStem or prepared in house) in knockout DMEM (Thermo Fisher Scientific) supplemented with 20% knockout serum replacement, 0.1 mM non-essential amino acids, 0.1 mM 2-mercaptoethanol, 100 U ml^{-1} penicillin, 0.1 mg ml^{-1} streptomycin, 2 mM L-glutamine, 20 ng ml^{-1} human LIF (Stem Cell Institute, University of Cambridge (SCI)), 8 ng ml^{-1} bFGF (SCI), 1 ng ml^{-1} TGFβ (Peprotech), 3 µM GSK3i (CHIR99021, Miltenyi Biotec), 1 µM ERKi (PD0325901, Miltenyi Biotec), 5 µM p38i (SB203580, TOCRIS Bioscience) and 5 µM JNKi (SP600125, TOCRIS Bioscience), as reported[14]. Cells were passaged every 2–4 days using TrypLE Express (Thermo Fisher Scientific). Before seeding 4i ES cells on MEFs, 10 µM of ROCKi (Y-27632, TOCRIS Bioscience) was added into the medium. Conventional ES cells were maintained on vitronectin (Thermo Fisher Scientific)-coated plates in Essential 8 medium (Thermo Fisher Scientific) according to the manufacturer's protocol. Cells were passaged every 3–5 days using 0.5 mM ethylenediaminetetraacetic acid (EDTA)/PBS.

To induce PGCLCs, 4i ES cells or preME (see below) cells were trypsinized into single cells and seeded into Corning Costar Ultra-Low attachment multiwell 96-well plates (Sigma) or AggreWell Microwell Plates (Stemcell Technologies) at 4,000–8,000 cells per well. PGCLC induction medium based on aRB medium contains 500 ng ml^{-1} BMP2 (SCI), 100 ng ml^{-1} SCF (Peprotech), 50 ng ml^{-1} EGF (R&D Systems) and 10 µM ROCKi. aRB medium is composed of Advanced RPMI 1640 Medium (Thermo Fisher Scientific) supplemented with 1% B27 supplement (Thermo Fisher Scientific), 0.1 mM non-essential amino acids, 100 U ml^{-1} penicillin–0.1 mg ml^{-1} streptomycin and 2 mM L-glutamine[13]. For DM^{+}PGCLC induction, PGCLC induction medium was replaced with aRB medium containing 100 ng ml^{-1} ActA (SCI), 20 µM Ra (Sigma), 100 ng ml^{-1} SCF (Peprotech) and 50 ng ml^{-1} EGF (R&D Systems) as indicated. For preME induction, trypsinized ES cells cultured in E8 were seeded on vitronectin-coated dish at 200,000 cells per well in 12-well plates in preME induction medium that is composed of aRB medium supplemented with 100 ng ml^{-1} ActA (SCI), 3 µM GSK3i and 10 µM ROCKi. For induction of exogenous transgenes, 100 µM DEX (Sigma) and/or 1 µg ml^{-1} dox (Sigma) was added.

Vector construction and transfection

For construction of reporter knock-in targeting vector, 5′ and 3′ arms amplified from human genomic DNA and combined with tdTomato or mVenus and Rox–PGK–PuroΔtk–Rox were cloned into modified NANOS3–tdTomato targeting vector containing MC1-promoter-driven diphtheria toxin A using in-fusion HD cloning kit (Takara Bio)[13]. Guide RNAs targeting around the stop codon sequence of DMRT1 or DAZL genes (Supplementary Table 1) were cloned into pX330 (Addgene). For construction of dox-inducible system, DMRT1 and BCL2L1 were cloned into PiggyBAC pCMV–Tet3G vector used previously[13]. All fragments were amplified by PCR using PrimeSTAR MAX, PrimeSTAR GXL DNA polymerase (Takara Bio) or Q5 High-Fidelity DNA Polymerase (NEB) according to the manufacturer's protocol.

Plasmid transfection for gene targeting or transgene introduction was carried out with electroporation or lipofection as described before[13,14]. In brief, electroporation was carried out using Gene Pulser equipment (Bio-Rad) with 1–5 × 10^{6} 4i ES cells mixed with targeting vector and pX330 plasmid containing guide RNA. For lipofection, reverse transfection was carried out with 2 × 10^{5} 4i ES cells in 100–200 µl of Opti-MEM containing plasmid vectors and Lipofectamine 2000 or Lipofectamine Stem Transfection Reagent (Thermo Fisher) with 5 min incubation at room temperature. After electroporation or lipofection, ES cells were seeded onto 4 drug resistant (DR4) MEFs (GlobalStem or SCI) and 48 h later, 0.5 µg ml^{-1} puromycin (Sigma) or 25 µg ml^{-1} hygromycin B (Thermo Fisher Scientific) was added to the culture medium for selection. Drug-resistant ES cell colonies were picked up and genotyped for correct targeting by PCR using primers in Supplementary Table 1. The targeted clones were expanded and then used for excision of Rox-flanked PGK–PuroΔtk by transient transfection of pCAG–Dre–IH. After selection with 25 µg ml^{-1} hygromycin B and subsequently with 0.2 µM fialuridine, colonies were picked up and assessed for excision by PCR using primers in Supplementary Table 1 (Extended Data Fig. 1c).

qPCR

Total RNA was extracted using PicoPure RNA Isolation Kit (Thermo Fisher) and cDNA was synthesized using QuantiTect Reverse Transcription Kit (QIAGEN) according to manufacturer's protocols. RT–qPCR was performed using QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher). Primer sequences are listed in Supplementary Table 3. Values shown were normalized to housekeeping genes and relative changes to control sample values.

Genomic DNA was extracted using Quick-DNA Microprep Plus Kit (Zymo). Primer sequences for genomic DNA quantification are listed in Supplementary Table 4. Values shown were normalized to

human genomic locus for TPOX and normalized to wild-type sample copy numbers.

Immunofluorescence and image analysis

Aggregates were fixed in 4% paraformaldehyde for 1–2 h at 4 °C and embedded in OCT compound (VWR) for frozen sections. Sections were incubated with primary antibodies for 1–2 h at room temperature or overnight at 4 °C and with fluorescent-conjugated secondary antibodies (dilution 1:500) for 1 h at room temperature. Primary antibodies are listed in Supplementary Table 5 (anti-DMRT1, rabbit, monoclonal, Abcam, cat. no. ab166893, dilution 1:500; anti-POU5F1, mouse, monoclonal, BD Biosciences, cat. no. 611203, dilution 1:500; anti-DAZL, rabbit, polyclonal, Abcam, cat. no. ab34139, dilution 1:200; anti-5mC, rabbit, monoclonal, Cell Signaling Technology, cat. no. 28692, dilution 1:200; anti-5mC, mouse, monoclonal, Abcam, cat. no. ab10805, dilution 1:150; anti-5hmC, rabbit, polyclonal, active motif, cat. no. 39769, dilution 1:500; anti-DNMT3B, sheep, polyclonal, R&D Systems, cat. no. AF7646, dilution 1:200; anti-TFAP2C, rabbit, polyclonal, Santa Cruz Biotechnology, cat. no. sc-8977, dilution 1:200; anti-SOX9, goat, polyclonal, R&D Systems, cat. no. AF3075-SP, dilution 1:200; anti-tdTomato, goat, polyclonal, SICGEN, cat. no. AB8181, dilution 1:100; anti-DDX4, rabbit, monoclonal, Abcam, cat. no. 235442, dilution 1:200; anti-mitochondria, mouse, monoclonal, Abcam, cat. no. ab92824, dilution 1:800; anti-SOX17, goat, polyclonal, R&D Systems, cat. no. AF1924, dilution 1:100; APC conjugated SUSD2, mouse, monoclonal, BioLegend, cat. no. 327408, dilution 1:100; anti-TFCP2L1, goat, polyclonal, R&D Systems, cat. no. AF5726, dilution 1:100). After antibody treatment, sections were stained with 4′,6-diamidino-2-phenylindole (Sigma) and imaged using Leica SP8 inverted laser scanning confocal microscope by white laser. HC PL APO CS2 63× 1.4 numerical aperture oil immersion objective was used. Image analyses were performed using a custom script[92] for Fiji[93], which segments nuclei in 4′,6-diamidino-2-phenylindole channel with difference of Gaussian threshold using Otsu's method[94] and measures intensity in channels for 5mC, 5hmC.

Flow cytometry analysis

Aggregates were trypsinized with trypsin/EDTA (0.25%, Thermo Fisher) at 37 °C for 5–15 min and single-cell suspension was incubated with Alexa Fluor 488 or 647 conjugated anti-alkaline phosphatase (TNAP) antibody (BD Bioscience, 5 µl per sample), PerCP–Cy5.5-conjugated anti-CDH5 antibody (BioLegend, 5 µl per sample) and/or Alexa Fluor 647 conjugated anti-CD38 antibody (BioLegend, 5 µl per sample) and analysed using BD LSRFortessa Cell Analyzer (BD Bioscience). Flow cytometry data were analysed using FlowJo software.

Luciferase assay

For vector construction of luciferase assay, three genomic regions with DMRT1 binding peaks containing DMRT1 motif (hg38; peak 1: chr3:16,608,590–16,608,949, DMRT1 motif: aaaactatgttact; peak 2: chr3:16,602,880–16,603,116, DMRT1 motif: aatacatagtagta; peak 3: chr3:16,594,400–16,597,625 DMRT1 motif: ttgatacaatgttt) in day 4 DZ^+PGCLCs at DAZL locus were amplified from human genomic DNA. These sequences were cloned into a piggyBAC-based luciferase (Luc+) reporter plasmid containing a hygromycin-resistant gene driven by a PGK promoter using in-fusion HD cloning kit. DMRT1 motif is scanned using HOMER scanMotifGenomeWide.pl function. The sequences without DMRT1 motif were amplified from the original plasmid with each peak's sequences using the primers listed in Supplementary Table 2. ALR/alpha consensus sequences (aattctcagtaacttccttgtgttgtgtgtattcaactcacagagttgaacgatcctttacacagagcagacttgaaacactctttttgtggaatttgcaagtggagatttcagccgctttgaggtcaatggtagaataggaaatatcttcctatagaaactagacagaat, DMRT1 motif sequence: ttgaaacactctttt) were downloaded from Repbase. The synthesized ALR oligos from Merck were cloned into a piggyBAC-based luciferase (Luc+) reporter plasmid containing a hygromycin-resistant gene driven by a PGK promoter using in-fusion HD cloning kit.

HEK 293 cells (ATCC CRL-1573) were transfected using Lipofectamine 2000 Transfection Reagent (Thermo Fisher) with a piggyBAC plasmid containing a constitutively expressed green fluorescent protein (GFP) cassette and a neomycin-resistant cassette, a piggyBAC plasmid containing a dox-inducible DMRT1 transgene and a puromycin-resistant cassette, and a plasmid encoding a piggyBAC transposase. Following 4 days of ±dox treatment, cells were measured for GFP with Hidex Sense (HIDEX) and subjected to luciferase activity assay using the Dual-Glo Luciferase Assay System (Promega). Normalized luciferase activities were obtained by dividing firefly luciferase activity by GFP signal counts.

Western blot

Nuclear proteins were extracted using EpiQuik Nuclear Extraction Kit II (EPIGENTEK) and were separated on a Novex 4–20% Tris-Glycine Mini Gel (Thermo Fisher) using XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher) and transferred to Hybond P 0.45 µm polyvinylidene fluoride membrane (GE Healthcare). After blocking in 5% skimmed milk, the membrane was incubated with primary antibodies (anti-SOX17, rabbit, monoclonal, Cell Signaling Technology, cat. no. 81778, dilution 1:1,000; anti-PRDM1, rabbit, monoclonal, Cell Signaling Technology, cat. no. 9115, dilution 1:500; anti-DMRT1, rabbit, monoclonal, Abcam, cat. no. ab126741, dilution 1:1,000; anti-LaminB1, rabbit, polyclonal, Abcam, cat. no. ab16048, dilution 1:1,000; Supplementary Table 6). The antibody binding was detected by horseradish-peroxidase-conjugated anti-rabbit IgG (Dako; dilution 1:2,000 in 0.01% TBST) in conjunction with the Western Detection System (GE Healthcare).

Preparation of scRNA-seq libraries

Reporter or cell surface marker-positive cells were sorted by BD FACSAria III Cell Sorter and loaded according to the manufacturer's protocol for the Chromium Next GEM Single Cell 3′ Reagent Kits v3.1 (Dual Index) (10x Genomics) to attain between 2,000 and 6,000 cells per reaction. Library preparation was carried out according to the manufacturer's protocol. Libraries were sequenced, aiming at a minimum coverage of 40,000 raw reads per cell, on the Novaseq 6000 systems using the sequencing format: read 1, 28 cycles; i7 index, 10 cycles; i5 index, 10 cycles; read 2, 90 cycles.

Preparation of bulk RNA-seq libraries

RNA-seq library was generated with total RNA (300 ng) using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (E7760, NEB) with NEBNext rRNA Depletion Kit v2 (NEB) according to manufacturer's protocol. Library was quantified using NEBNext Library Quant Kit Quick Protocol (E7630, NEB). Libraries were sequenced for 150 cycles in paired-end mode on the NovaSeq platform.

C&R

C&R for DMRT1 and normal rabbit IgG was performed as described[43–45]. Briefly, 50,000 purified DZ^+PGCLCs were washed and bound to activated 10 µl Concanavalin A-coated magnetic beads. The beads were then incubated with wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine and protease inhibitor) containing 0.1% digitonin and 1 µg of DMRT1 antibody (ab126741, Abcam) or normal rabbit IgG (#2729, Cell signaling) for 2 h at 4 °C on a rotator. After two washes in digitonin–wash buffer, beads were resuspended in Protein A/G-MNase fusion protein at 70 ng ml^{-1} in digitonin–wash buffer and incubated for 1 h at 4 °C on a rotator. After two washes in digitonin–wash buffer (the beads with replicate 3 of day 4 DZ^+PGCLC and day 8 DZ^+PGCLC were washed with low-salt rinse buffer (20 mM HEPES, pH 7.5, 0.5 mM spermidine and 0.1% digitonin) once additionally), beads were resuspended in ice-cold calcium incubation buffer (3.5 mM HEPES pH 7.5,

10 mM $CaCl_2$ and 0.1% digitonin). After 15 min, 2× stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM egtazic acid, 0.1% digitonin, RNase A 100 µl ml^{-1} and glycogen 50 µg ml^{-1}) was added. Beads were incubated at 37 °C for 30 min, the liquid was removed to a fresh tube and DNA was extracted with phenol–chloroform extraction.

DNA library preparation and sequencing

Sequencing libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit (NEB, E7645S) for Illumina according to the manufacturer's protocol but without size selection and PCR enrichment of adaptor-ligated DNA. PCR enrichment of adaptor-ligated DNA was performed with KAPA HiFi Real-Time PCR Library Amplification Kit (Roche, KK2702) following the manufacturer's recommendations. The number of PCR cycles using the KAPA polymerase was 7–10. SPRIselect beads (Beckman Coulter, B23317) were used for clean-up PCR product and size selection. Libraries were sequenced for 150 cycles in paired-end mode on the NovaSeq platform.

TAPS with βGT blocking and chemical-assisted pyridine borane sequencing plus

TAPS with βGT blocking (TAPSβ) and chemical-assisted pyridine borane sequencing plus (CAPS+) were performed according to previous publications[47,48]. Briefly, DNA was spiked with spike-in control DNA and sonicated to 300–500 bp, before ligation with NEBNext Adaptor for Illumina using KAPA HyperPrep Kit according to the manufacturer's protocol. The uracil in the loop of NEBNext Adaptor was removed by USER Enzyme (New England Biolabs). A total of 100 ng ligated DNA was used for both TAPSβ and CAPS+. For TAPSβ, the ligated library was subjected to βGT (Thermo Fisher) blocking, two rounds of mTet1 oxidation, and borane reduction. For CAPS+, the ligated library was subjected to chemical oxidation and borane reduction. Converted DNA from TAPSβ and CAPS+ was amplified with NEBNext Multiplex Oligos for Illumina and KAPA HiFi HotStart Uracil+ ReadyMix PCR Kit for four cycles according to the manufacturer's protocol. The PCR product was purified with Ampure XP beads. Libraries were sequenced for 150 cycles in paired-end mode on the NovaSeq 6000 platform.

Data processing for scRNA-seq

The reads were demultiplexed and aligned to the 10x Genomics' GRCh38-2020-A reference genomes using the Cell Ranger Software (v.7.0.0, 10x Genomics) with default parameters. The summary statistics from Cell Ranger is provided in Supplementary Table 7.

We employed Scrublet to identify and distinguish single cells from cell doublets in each individual library. As described in ref. 95, we used a two-step diffusion doublet identification followed by Bonferroni–false discovery rate (FDR) correction and a significance threshold of 0.01. We used Scanpy v.1.8.0 (ref. 96) to analyse the filtered count matrices that were generated by Cell Ranger, following their recommended standard practices. Specifically, we excluded genes that were expressed by fewer than three cells and excluded cells that expressed fewer than 3,000 genes or had more than 10% mitochondrial content. We then normalized the raw counts by library size and log-transformed them. Next, we identified the highly variable genes, which we used for principal components analysis (PCA). We corrected for the library effect using Harmony[97] on the PCA space (default parameters except theta = 1). Finally, we used the Harmony-corrected PCA space to identify the k (k = 15) nearest neighbours, perform Leiden clustering and visualize the results using UMAP. Leiden clusters with overall high doublet score or low counts number were flagged and discarded in further analysis. We used Seurat's v.4.0.5 FindAllMarkers() function to identify up- and downregulated genes in each library with |$\log_2$fold change (FC)| >1 (ref. 98). To determine the cell cycle phase (that is, G1, S or G2/M) of each cell, we combined the expression of G2/M and S phase markers and used the method implemented in Scanpy's score_genes_cell_cycle function to classify the cells[99]. We then compared the in vitro cell states identified in our study with the in vivo cell states reported in the Smart-seq2 dataset of gonadal cells from Li et al.[24] (GSE86146). To do this, we downloaded the normalized transcripts per million (TPM) matrix from Li et al.[24] and annotated their cells using the 'FullAnnot' field. We only considered the male foetal germ cell clusters. We used the tool scmap[100] to project the Li et al.[24] annotations onto our dataset and visualized the results of the projections using a dot plot.

Data processing for bulk RNA-seq

Trim Galore[101] was used to remove the low-quality reads and adaptor sequences. Trimmed sequence files were mapped to human reference genome (GENCODE, GRCh38.p13) and counts on genes were generated using STAR[102] with parameters –outFilterMultimapNmax 1 –outFilterMatchNmin 35. Normalized counts (normalize the total number of mapped reads per experiment to 1×10^8) on repeat elements were generated with the analyzeRepeats.pl of the HOMER[103] package. Differential gene (or repeat element) expression analysis was performed with the glm method of the edgeR[104] package for protein-coding genes. DEGs or repeat elements were identified with fold changes greater than 2 and FDR smaller than 0.05. Reads per kilobase of transcript per million mapped reads (RPKM) values of genes were calculated using Cufflinks[105].

Secondary data analyses were performed using Microsoft Excel and R software version 4.0.5 with the packages ggplot2. GSEA[106] was performed using the GSEA software by the Broad Institute. GO analysis was performed on the basis of GO Biological Process (http://geneontology.org). Marker protein-coding genes, 142 for migratory, 288 for mitotic and 937 for mitotic arrest male PGCs, were used on the basis of published markers identified from single cell RNA-seq data[24,28], 'PGC genes' were identified on the basis of shared enriched DEGs (logFC >1, FDR <0.05) between week 7 and week 9 male PGCs against week 7 gonadal somatic cells or conventional ES cells[6].

Data processing for C&R

To trim the short fragments that are frequently encountered in C&R experiments we used leeHom package program[107] with –ancientdna option. The trimmed reads were aligned to the human reference genome (GENCODE, GRCh38.p13) using Bowtie2 2.2.6 (ref. 108) with options –very-sensitive –no-mixed –no-discordant -q –phred33 -I 10 -X 700. For MACS2 peak calling, parameters used were macs2 (ref. 109) callpeak –keep-dup all and the peaks with $-\log_{10}(q \text{ value})$ >10 for day 4 DZ^+PGCLC and the peaks with $-\log_{10}(q \text{ value})$ >8 with IgG as control for day 8 DZ^+PGCLC were selected. A total of 11,920 (day 4) and 7,818 (day 8) peaks that are in common between the replicates were used for further analysis. Peaks were annotated to their nearest genes or overlapping repeat elements using Homer annotatePeaks.pl function. To analyse the enriched TF motifs over peaks or repeat elements, HOMER findMotifsGenome.pl function was used.

Data processing for TAPSβ and CAPS+ methylome

The reads were demultiplexed using i7 sequences. The total sequencing reads number and conversion rate are provided in Supplementary Table 8. Trim Galore was used to remove the low-quality reads, and Samtools rmdup function was used to remove PCR duplicates. Trimmed reads were mapped to human reference genome (GENCODE, GRCh38.p13), and modified bases were called by asTair[110]. The methylation rate (%) for each CpG was calculated as the ratio between T and (C + T). Average CpG methylation levels of annotated genomic regions were calculated using UCSC bigWigAverageOverBed considering only information from CpGs with >5× coverage. To identify DMRs, we used DMRfinder[111] with the default setting except –meanDiff_cutoff (5mC, 0.2; 5hmC, 0.05) and –pctMinCtrl 0 –pctMinExp 0 as sets of CpGs with a t-statistic greater than the critical value for α = 0.05 and with a gap <300 bases.

Statistics and reproducibility

For RNA-seq, C&R and 5hmC/5mC methylome data, two independent biological replicates (except for day 4 DZ^{+}PGCLC C&R with three independent biological replicates) were included according to the guidelines of the Encode Consortium101. No statistical method was used to pre-determine sample size in other experiments. Low-quality replicates of libraries were excluded from the analysis, as determined by percentage of reads in peaks, number of peaks and genome browser visualization. As all results involved equipment-based quantitative measure and no subjective rating of data was involved, blinding and randomization are not relevant. All the data met the assumptions of the statistical tests used, including whether normality and equal variances were formally tested. All the data collection and analysis were not performed blind to the conditions of the experiments.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE223036. Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Code availability

The script we used for the image analysis is released under the GPL v3, which is included in the GitHub repository (https://github.com/gurdon-institute/Nucleus_Measure/blob/main/LICENSE) and allows anyone to use, modify and distribute the software under the same license without warranty or liability.

References

90. Bryja, J. & Konecny, A. Fast sex identification in wild mammals using PCR amplification of the Sry gene. *Folia Zool.* **3**, 269–274 (2003).
91. Gafni, O. et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
92. Nucleus_Measure. *GitHub* https://github.com/gurdon-institute/Nucleus_Measure/blob/main/Nucleus_Measure.py (2021).
93. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
94. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cyber.* **9**, 62–66 (1979).
95. Popescu, D.-M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
96. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
97. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
98. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
99. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
100. Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15**, 359–362 (2018).
101. Trim Galore. *Babraham Bioinformatics* https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (2012).
102. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
103. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
104. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
105. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
106. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
107. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141–e141 (2014).
108. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
109. Zhang, Y. et al. Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9**, R137 (2008).
110. asTair. *Bitbucket* https://bitbucket.org/bsblabludwig/astair/src/master/ (2019).
111. DMRfinder. *GitHub* https://github.com/cemordaunt/DMRfinder (2018).

Acknowledgements

We thank J. Cerviera, M. Jarana and C. Bradford for assistance with flow cytometry and cell sorting; R. Butler for generating a program for immunofluorescence image analysis; S. Kim for critical suggestions for the study and technical assistance; the Austin Smith lab for providing Shef-6 cell line and technical discussions; the Rick Livesey lab for supporting tissue culture; H. Saito, M. Gu, N. Merleau-Ponty, E. Sen and C. Lee for technical assistance; former and current Surani lab members and Gurdon Institute members for input, discussion and support for the research; and M. Suematsu, the Department of Biochemistry & Integrative Medical Biology, School of Medicine, Keio University and C. le Sage for critical support. The research was supported by an MRC research grant (RG85305) and Rosetrees Trust to M.A.S. and N.I., by a Wellcome Investigator Awards in Science (209475/Z/17/Z and 096738/Z/11/Z) and a BBSRC research grant (G103986) to M.A.S., by JSPS KAKENHI Grant (JP23H03047) to N.I. and by a core grant to the Gurdon Institute by Wellcome and Cancer Research UK. S.-M.L. is supported by the Brain Pool programme funded by the Ministry of Science and ICT through the National Research Foundation of Korea (2022H1D3A2A02063272). T.K. is supported by AMED (JP22bm1123008). The C.-X.S. lab is supported by the Ludwig Institute for Cancer Research, Cancer Research UK (C63763/A26394 and C63763/A27122), and National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). H.X. is supported by China Scholarship Council. M.I. is supported by the Nakajima Foundation. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Author contributions

The study was conceived and designed by N.I. and M.A.S. N.I. designed research strategies and experiments, and performed cell culture, molecular experiments and data analysis. S.-M.L. designed research strategies and experiments, and performed molecular experiments and bioinformatics. E.D. provided technical support for cell culture and molecular experiments. T.K. provided technical advances and materials. V.L., C.S.-S and R.V.-T performed and analysed single-cell RNA-seq. H.X., J.C., M.I. and C.-X.S. performed methylome sequencing. S.D. performed bioinformatics. N.I., M.A.S. and S.-M.L. wrote the manuscript with inputs from all authors.

Competing interests
The authors declare no competing interests.

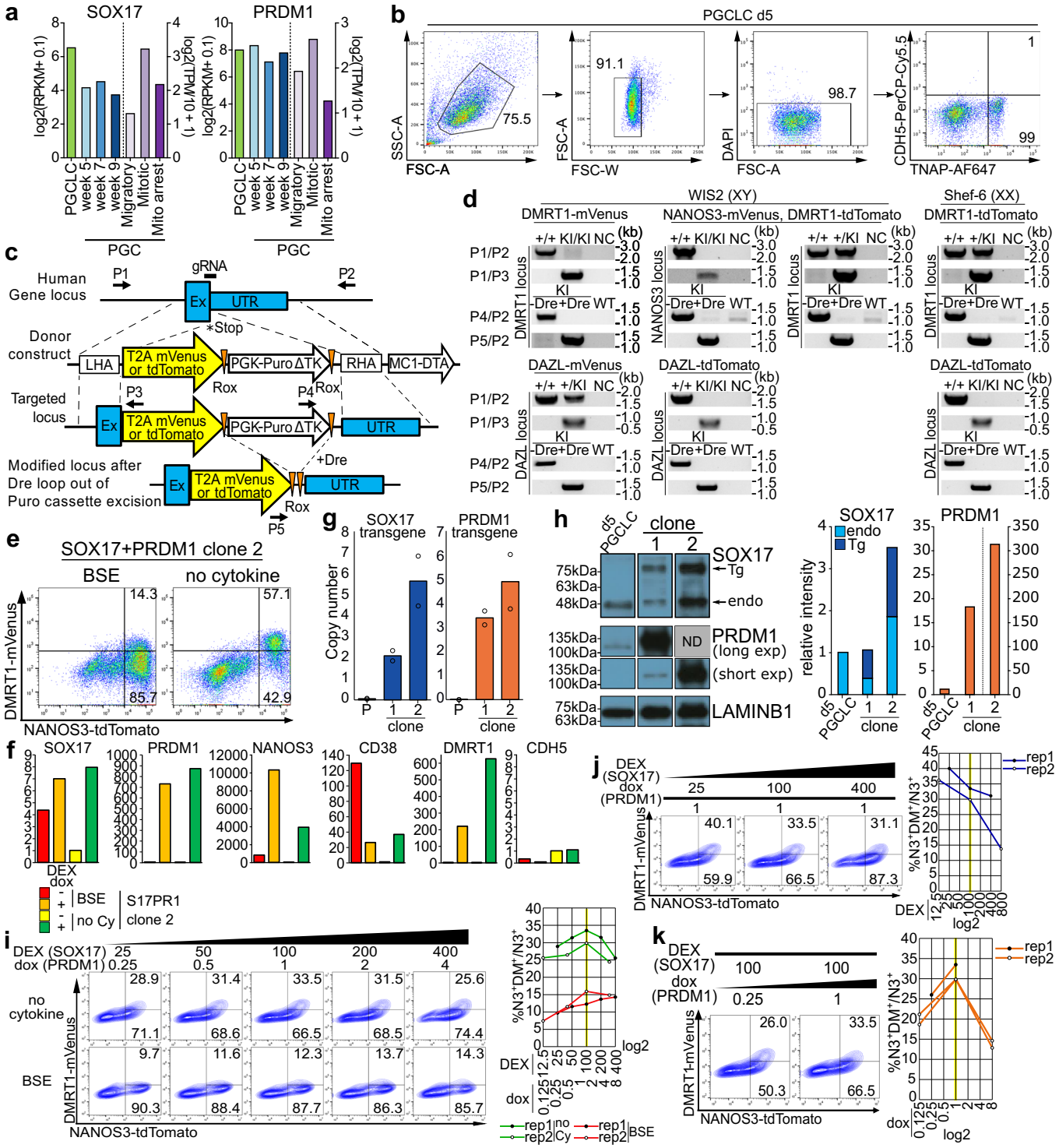
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41556-023-01224-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-023-01224-7.

Correspondence and requests for materials should be addressed to Naoko Irie or M. Azim Surani.

Peer review information *Nature Cell Biology* thanks Jie Qiao, Jinlian Hua and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41556-023-01224-7

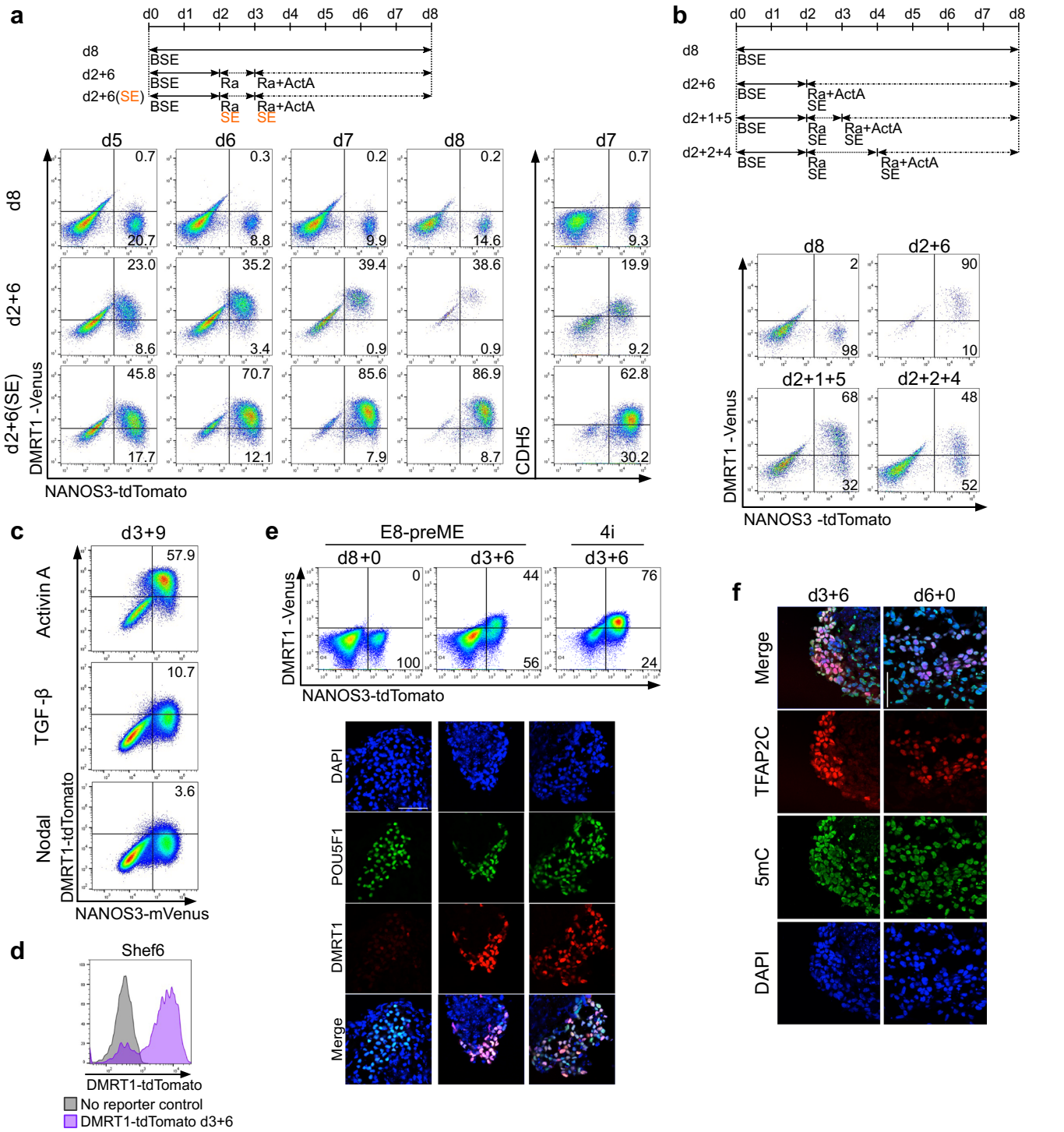


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41556-023-01224-7

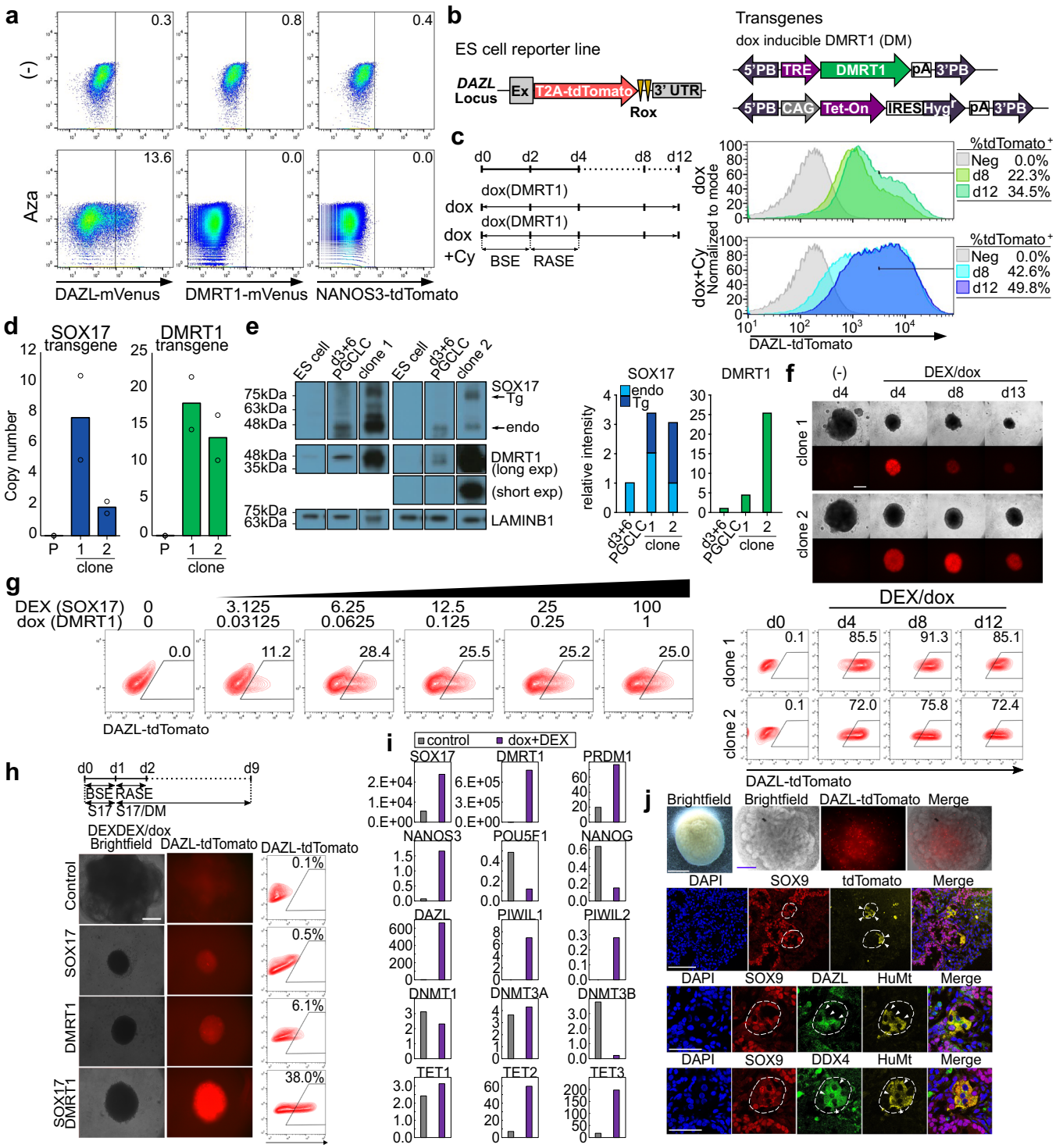
Extended Data Fig. 1 | Sustained signaling for PGCLC specification restricts subsequent development. (**a**) RNA-sequencing data 6,14,24 of in vitro induced day 4 PGCLCs and in vivo PGCs for PGC genes. Bar plot represents mean values (PGC week5/7/9, n = 2 biological samples; migratory, 37 cells; mitotic, 332 cells; mitotic arrest, 309 cells). (**b**) Example of gating strategies for flowcytometry analysis with PGCLC d5 from Fig. 1b. Cell population is gated based on SSC-A/FSC-A, single cell population based on FSC-A/FSC-W and live cell population using DAPI. % gating populations/the parental populations. (**c**) Targeting strategy for reporter ES cells. Guide (g)RNA for CRISPR-Cas9 and genotyping primers (P1-P5) are indicated. (**d**) Genotyping PCR for reporter ES cells. +: wild type allele; KI: knock-in allele; NC: water control; WT: wildtype cells; −/+DRE: before (−) and after (+) puromycin resistant cassette excision. The experiment was repeated independently two times with similar results. (**e**) Flowcytometry and (**f**) RT-qPCR of PGCLC induction from NANOS3−tdTomato/DMRT1-mVenus double reporter male ES cells (WIS2) with inducible SOX17 = GR and TRE-PRDM1 cell line (clone 2) treated with DEX and dox (SOX17 + PRDM1) for five days in the presence or absence (no cytokine) of PGC specification cytokines, BMP2, SCF and EGF (BSE). Percentage for +/−DMRT1-mVenus in NANOS3-tdTomato positive cells are indicated. Values in (F) are normalized with the housekeeping gene GAPDH and relative changes against no tg/no Cy/DEXdox(-). A repeat experiment with independent clone with similar results is shown in Fig. 1e. (**g**) Quantitative PCR for copy number for genomic integration of the piggybac plasmids for SOX17 = GR and TRE-PRDM1 in clone 1 and 2. Biologically independent experiments, n = 2. (**h**) Western blot for expression of SOX17, PRDM1, LAMINB1 in the nuclear protein extract for the transgenic cell lines 1 and 2 in the presence of DEX and dox for 3 days, and PGCLCs for day 5 induced with cytokines, BMP2, SCF and EGF. The experiment was repeated independently two times with similar results. Dose response of DEX (μM) and dox (μg/ml) (**i**), DEX (**j**) and dox (**k**) for the expression of DMRT1 reporter in NANOS3+PGCLCs. Percentage for +/−DMRT1-mVenus in NANOS3-tdTomato positive cells are indicated. The graphs show the percentages from the two independent experiments.

https://doi.org/10.1038/s41556-023-01224-7



Extended Data Fig. 2 | Switching signaling in nascent PGCLCs for further development. (**a** and **b**) Flowcytometry of NANOS3-tdTomato against DMRT1-mVenus. NANOS3-tdTomato/DMRT1-mVenus reporter 4i ES cells were induced PGCLCs with the cytokines indicated for 8 days. BSE: BMP2, SCF and EGF; Ra: retinoic acid, 20 μM; ActA: Activin A, 20 μg/ml; SE: SCF and EGF. Percentage populations of NANOS3 positive and DMRT1 positive or negative cells in live single cells are shown. (**c**) Flowcytometry of NANOS3-tdTomato against DMRT1-mVenus. 3-days induction of nascent PGCLCs with BSE followed by Ra 20 μM, SCF 100 μg/ml, EGF 50 μg/ml and, Activin A 100 μg/ml, TGFβ 100 μg/ml or Nodal 100 μg/ml. Percentage populations of NANOS3 and DMRT1 double positive cells in live cell population are shown. (**d**) Flowcytometry histogram of DMRT1-tdTomato. DMRT1-tdTomato female reporter line (Shef-6) was induced nascent PGCLCs with BSE for 3 days and followed by Ra 20 μM+Activin A 100 μg/ml+SCF 100 μg/ml+EGF 50 μg/ml for 6 days (d3 + 6). Parental Shef-6 ES cells without reporter was used as a negative control. (**e**) Flowcytometry and immunofluorescence for PGCLCs induced with BSE (d8 + 0) or BSE 3 days followed by RASE (d3 + 6) from conventional ES cells cultured in E8 via precursors of mesendoderm (E8-preME) or 4i ES cells (4i). Values in flowcytometry plots indicate percentage of DMRT1 positive or negative in NANOS3-tdTomato positive cell populations. Scale bar in immunofluorescence image shows 50 μm. The experiment was repeated independently two times with similar results. (**f**) Immunofluorescence of TFAP2C and 5-methylcytosine (5mC) for PGCLCs induced with BSE (d6 + 0) or BSE 3 days followed by RASE (d3 + 6). Scale bar =50 μm. The experiment was repeated independently two times with similar results.

https://doi.org/10.1038/s41556-023-01224-7

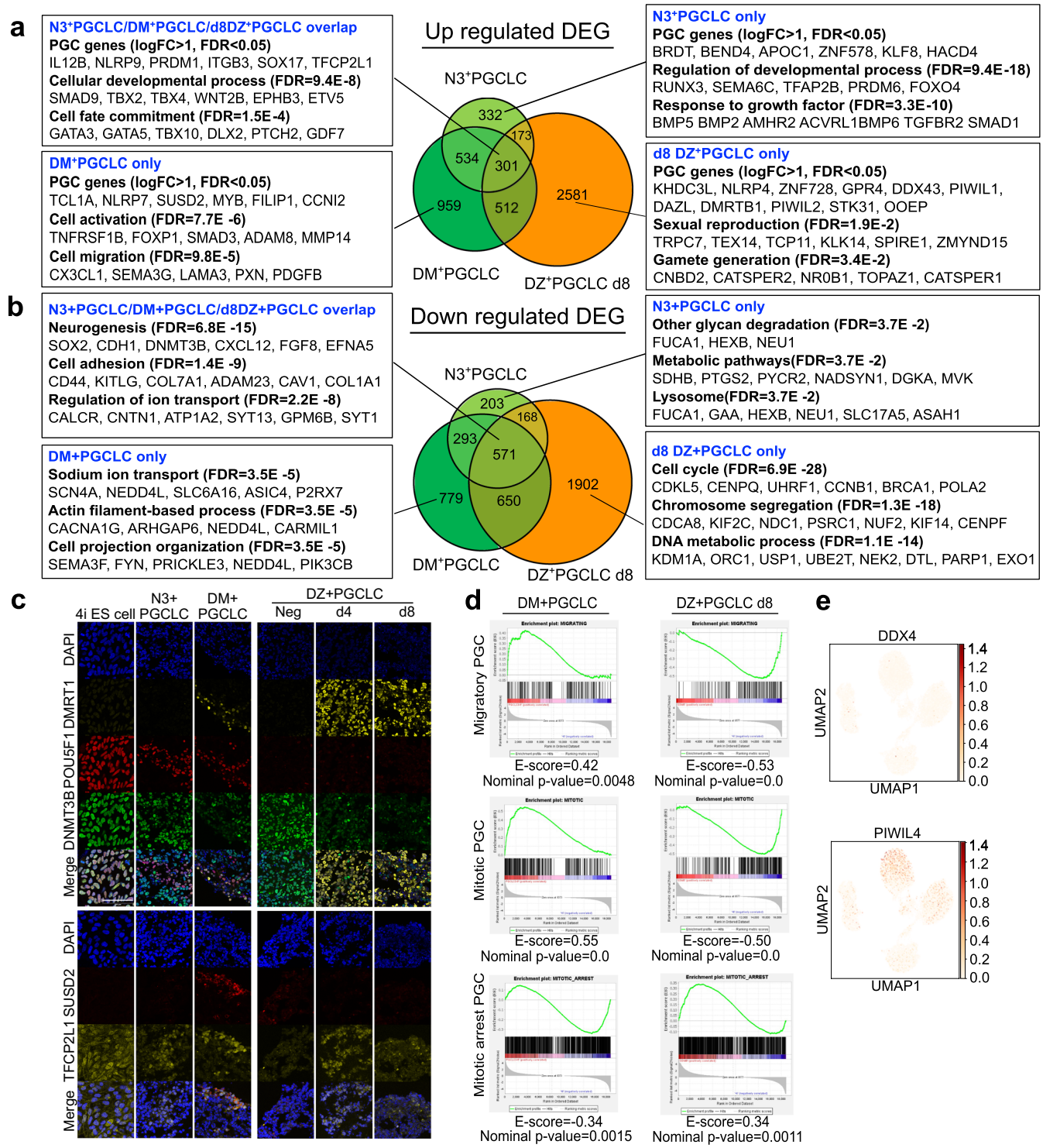


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41556-023-01224-7

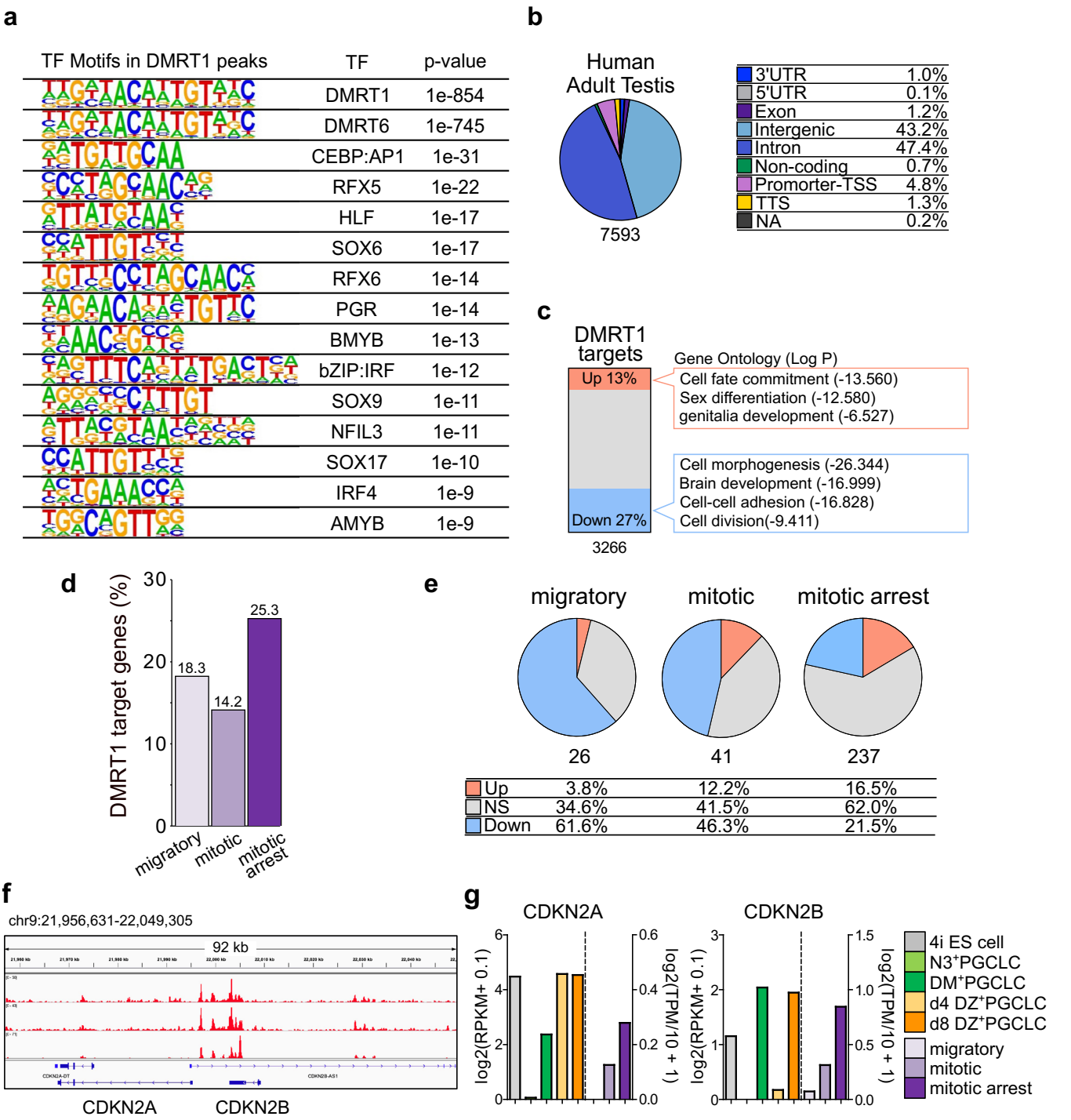
Extended Data Fig. 3 | DMRT1 activates DAZL expression. (**a**) Flowcytometry of DAZL-mVenus, DMRT1-mVenus and NANOS3-tdTomato ES cells treated with 5-Aza-2-deoxycytidine (Aza, 5 μM) for 3 days or untreated control (-). % reporter positive cells/live cell population. (**b**) Design for dox-inducible DMRT1 (DM) and DAZL–tdTomato ES cells. (**c**) Flowcytometry of DAZL-tdTomato ES cells with dox-inducible DMRT cultured for 8 or 12 days with doxycycline (dox) with or without cytokines (Cy): BSE followed by RASE. Negative controls (Neg) are Day 0 samples. %DAZL-tdTomato positive cells/live cell population. (**d**) Quantitative PCR for copy number of SOX17 = GR and TRE-DMRT1 plasmids in genomic DNA for clone 1 and 2 cell lines. Biologically independent experiments, n = 2. (**e**) Western blot for SOX17, DMRT1, LMNB1 in nuclear protein extract for transgenic cell lines treated with DEX/dox for 2 days and PGCLC d3 + 6. The experiment was repeated independently two times with similar results. (**f**) Fluorescence images and flowcytometry of DAZL-tdTomato WIS2 ES cells transgenic for clone 1 and 2 with (DEX/dox) or without (-) SOX17/DMRT1 induction treated with BSE followed by RASE as in Fig. 3b. Scale bar =200 μm. (**g**) Flowcytometry of DAZL reporter for dose response to DEX (μM) and dox (μg/ml). (**h**) Fluorescence images and flowcytometry of transgenic DAZL-tdTomato Shef-6 ES cells cultured for 9 days. Scale bar =200 μm. %DAZL-tdTomato positive cells/live cells. The experiment was repeated independently two times with similar results. (**i**) RT-qPCR of transgenic DAZL-tdTomato Shef-6 ES cells induced for 4 days with (dox+DEX) or without (control) dox/DEX. ΔΔCT normalized with RPLP0. The data represent from n = 1 experiment and repeat experiments with independent cell line are shown in Fig. 3c. (**j**) Day 11 testicular organoid reconstituted from E12.5 male mouse gonadal soma28 with DAZL-tdTomato positive PGCLC day 9 induced by DMRT1/SOX17. Brightfield (Top left, Scale bar = 200 μm) and fluorescence (Scale bar = 100 μm) microscopy images. Immunofluorescence of organoids for SOX9, tdTomato, DAZL, DDX4 and human mitochondria (HuMt). Dashed lines show testicular tubules with SOX9 expression. Arrowheads show tdTomato or HuMt positive human PGCLCs. Scale bars: 100 μm (top) and 50 μm (middle/bottom). The experiment was repeated independently two times with similar results.

https://doi.org/10.1038/s41556-023-01224-7



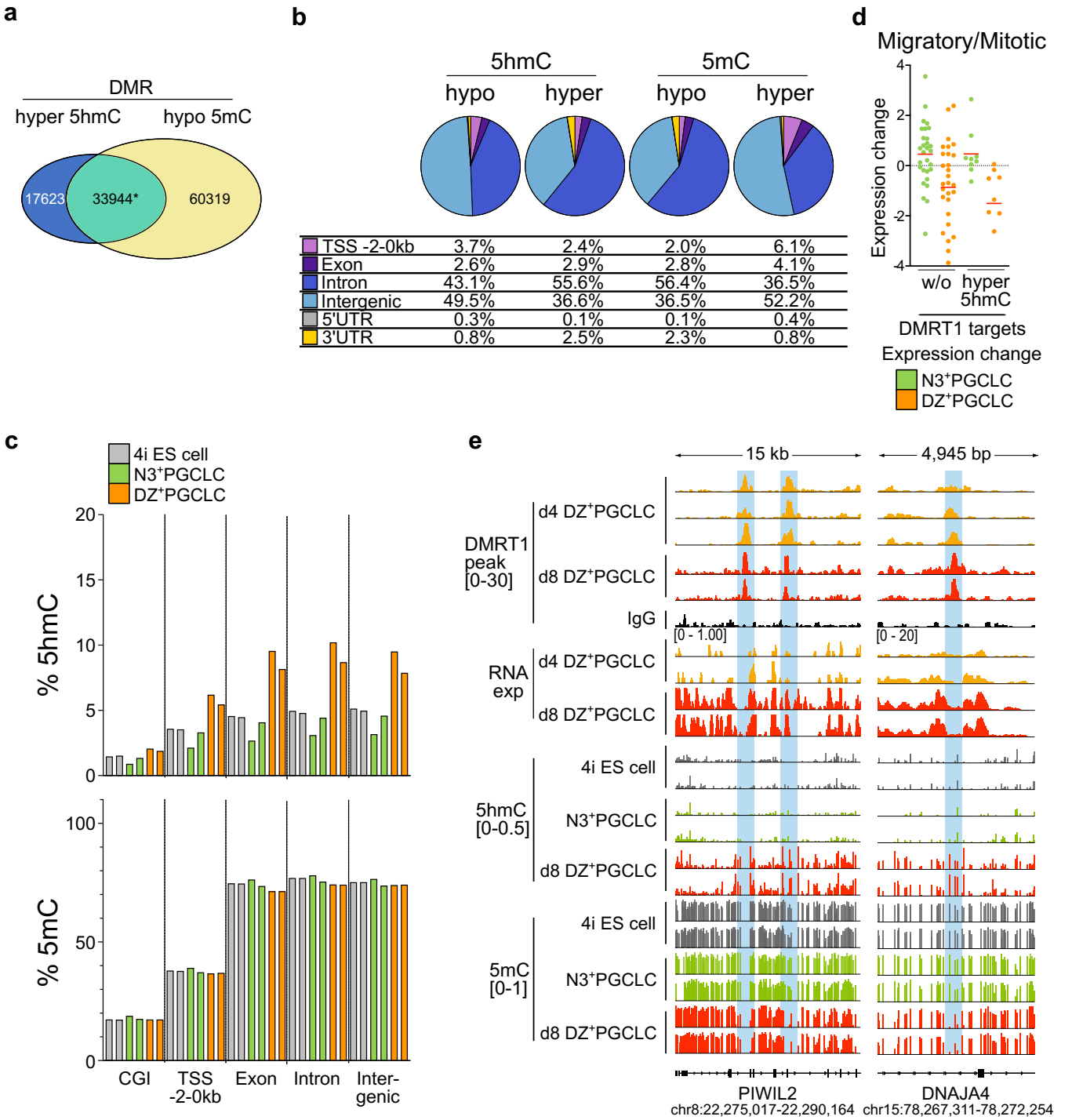
Extended Data Fig. 4 | Transcriptome network in in vitro induced PGCLCs. (A and B) Venn diagram for significantly upregulated (Log2 (FC) > 1 and FDR < 0.05) (**a**) and downregulated (Log2 (FC) <-1 and FDR < 0.05) (**b**) DEGs in N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC d8 versus 4i ES cells. The number of genes for each category is indicated in the Venn diagram. Text boxes show representative PGC genes and related Gene Ontology (GO) biological processes terms significantly enriched with FDR values. The data represents an integration of two biological replicates. FC: Fold Changes. (**c**) Immunofluorescence for marker genes in 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC (day 4 (d4), day 8 (d8), no DEX/dox control for 8 days (Neg)). TFCP2L1 for 4i ES cells shows high background signal. Scale bar: 100 µm. The experiment was repeated independently two times with similar results. (**d**) Gene set enrichment analysis (GSEA) for DM$^+$PGCLC and DZ$^+$PGCLC d8 versus 4i ES cells based on RNA-seq against marker genes for migratory, mitotic, and mitotic arrest PGCs. E-score (enrichment score) and nominal p-value are indicated. The p-values for the GSEA test statistics are calculated by permutation. The data represents an integration of two biological replicates. (**e**) UMAPs of 4i ES cells, N3$^+$PGCLC, DM$^+$PGCLC and DZ$^+$PGCLC d8 scRNA-seq dataset (each from single biological sampling points) batch-corrected with Harmony. Expression of *DDX4* and *PIWIL4* shown in log transformed (normalized count (NC) + 1).

https://doi.org/10.1038/s41556-023-01224-7



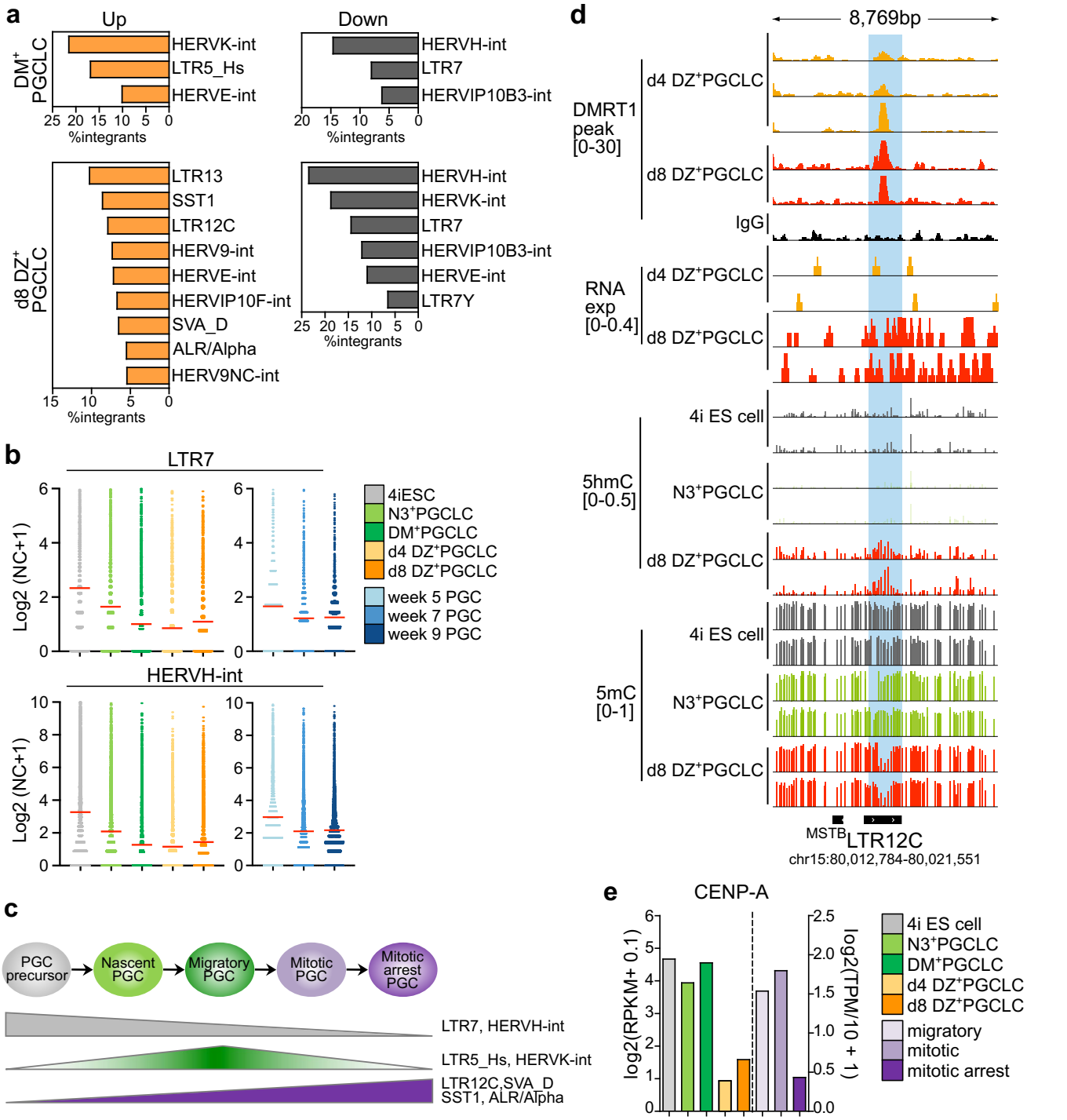
Extended Data Fig. 5 | DMRT1 targets for human fetal germ cell development. (**a**) Topmost commonly known transcription factor (TF) motifs enriched in all DMRT1 peaks in d4 DZ$^+$PGCLCs identified by C&R using Homer (cumulative binomial distributions). The data represents an integration of three biological replicates. (**b**) Distribution of genomic features with DMRT1 peaks identified in human adult testis[46]. Total number of peaks and proportion (%) of genomic features in all the peaks are indicated. (**c**) The bar plot represents the % of overlap between DMRT1 target genes and upregulated (Log2 Fold Changes (FC) > 1 and FDR < 0.05) and downregulated (Log2 (FC) < -1 and FDR < 0.05) genes in d8 DZ$^+$ PGCLC versus 4i ES cells. Text boxes show the related Gene Ontology (GO) biological processes terms significantly enriched. P-value is calculated based on a cumulative hypergeometric distribution. The data represents an integration of two biological replicates. (**d**) Proportion (y-axis, %) of DMRT1 targets within the marker genes for migratory (26/142 genes: 18%), mitotic (41/288 genes: 14%) or mitotic arrest PGCs (237/937 genes: 25%) in d8 DZ+PGCLCs. The data represents an integration of two biological replicates. (**e**) Proportion of DMRT1 targets for upregulated (Up, orange), not significant (NS, grey), and downregulated (Down, blue) genes for migratory, mitotic or mitotic arrest PGC in d8 DZ$^+$PGCLCs versus 4i ESCs. The data represents an integration of two biological replicates. (**f**) DMRT1 C&R peaks at CDKN2A and CDKN2B genomic locus in d4 DZ$^+$ PGCLC. Technical triplicates are shown. (**g**) Expression of CDKN2A and CDKN2B in 4i ES cells, PGCLCs and PGCs from RNA-seq. Y-axes are log2 (RPKM + 0.1), left and log2 (TPM/10 + 1), right. Bar plot represents the mean (4i ES cells and PGCLCs, n = 2 biological replicates; migratory, 37 cells; mitotic, 332 cells; mitotic arrest, 309 cells).

https://doi.org/10.1038/s41556-023-01224-7



Extended Data Fig. 6 | 5hmC and 5mC profiling in DZ+PGCLCs. (**a**) Venn diagrams showing the overlap of DMRs for hyper 5hmC and hypo 5mC in d8 DZ$^+$PGCLCs compared to 4i ES cells. Number of DMRs are indicated. *p-value = e-104768 for the fisher's extract test, two-tailed. The data represents an integration of two biological replicates. (**b**) Pie chart presents the distribution of genomic features for DMRs. The data represents an integration of two biological replicates. (**c**) Bar plots showing the methylation levels of different genomic features (CGI, n = 25612; TSS-2 ~ 0 kb, n = 26418; Exon, n = 207932; Intron, n = 149208; Intergenic, n = 17061). Two biological replicates of each cell type are shown. (**d**) Dot plot showing the expression changes (d8 DZ$^+$PGCLCs-4i ES cells) of marker genes for migratory/mitotic PGCs which have DMRT1 peaks with (n = 9 genes) or without (n = 29 genes) DMRs for hyper 5hmC in their intronic regions. The mean value is indicated by a red bar in the dot plot. The data represents an integration of two biological replicates. (e) DMRT1 C&R peaks, bulk RNA-seq, 5hmC and 5mC at PIWIL2 and DNAJA4 locus.

https://doi.org/10.1038/s41556-023-01224-7



Extended Data Fig. 7 | Regulation of genomic repeat elements and DMRT1 in human early germline. (**a**) Differentially expressed repeats at |log Fold Changes (FC)|>1 and FDR ≤ 0.05 in DM+ PGCLCs or d8 DZ+ PGCLCs versus 4i ES cells. The bar charts show the top subfamilies (> 200 total integrants in the human genome) with the percentage of differentially expressed integrants over total integrants. The data represents an integration of two biological replicates. (**b**) Dot plots for expression of genomic repeat subfamilies with read count≥1 in the sample set. Y-axis is log2 transformed (normalized count (NC) + 1). The mean value is indicated by a red bar in the dot plot. The data represents an integration of two biological replicates. (**c**) Schematic model for expression dynamics of repeat elements during human PGC development. (**d**) DMRT1 C&R peaks, bulk RNA-seq, 5hmC and 5mC at LTR12C in chromosome 15. (**e**) Expression of CENPA in 4iES cells, PGCLCs and PGCs from RNA-seq. Y-axes are log2 (RPKM + 0.1), left and log2 (TPM/10 + 1), right. Bar plot represents the mean (4iES cells and PGCLCs, n = 2 biological replicates; migratory, 37 cells; mitotic, 332 cells; mitotic arrest, 309 cells).

nature portfolio

Corresponding author(s): Naoko Irie, M. Azim Surani

Last updated by author(s): 17 July 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	-FACS was performed using BD LSRFortessa Cell Analyzer (BD Bioscience) -qPCR was performed on a QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher) -Immunofluorescences were performed using Leica SP8 inverted laser scanning confocal microscope -All libraries were sequenced using NovaSeq 6000 (Illumina)
Data analysis	-Flow cytometry data was analysed using FlowJo (10.7.1) -Immunofluorescences data was analysed using Fiji software (2.9.0) and a custom script (https://github.com/gurdon-institute/Nucleus_Measure/blob/main/Nucleus_Measure.py) -For quantification in qPCR analysis, The ΔCt or ΔΔCt method was used. -Bioinformatic software: Trim Galore (0.6.6), STAR (2.5.4b), cufflinks (2.2.1), R (4.0.5), edgeR (3.36.0), leeHom (1.2.5), asTair (3.3.2), cutadapt (1.15), Bowtie 2 (2.2.6), samtools (1.7), MACS2 (2.1.2), HOMER (4.11.1), bedtools (2.26.0), Seurat (v4.0.5), Cell Ranger (7.0.0), Scanpy (1.8.0), IGV (2.15.1), GSEA (4.1.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Bulk RNA sequencing datasets, single cell sequencing datasets, methylome and CUT&RUN sequencing data sets have been deposited in NCBI GEO with the accession code GSE223036.
Genome databases used are: Gencode Human Release GRC38.p13 and 10X Genomics GRCh38-2020-A. Source data are provided with this study. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For NGS libraries, two or three biological replicates were generated according to the practice of the Encode Consortium. For other experiments, no sample size calculation was performed. As per standard practice in molecular and cell biology, at least 2 replicates were generated.
Data exclusions	Low quality replicate of CUT&RUN libraies were excluded from the analysis.
Replication	The experimental findings were reliably reproduced with at least two or three replications.
Randomization	In each experiment, cells started from the same conditions and treatments were randomly allocated to experimental groups.
Blinding	All results involved equipment-based quantitative measure and no subjective rating of data was involved, hence blinding is not relevant.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

The primary antibodies used for immunofluorescence:
1. anti-DMRT1, Rabbit, Monoclonal, Abcam, Cat. number ab166893, Lot number GR119578-4 (1:500)
2. anti-POU5F1, Mouse, Monoclonal, BD Biosciences, Cat. number 611203, Lot number 8087969 (1:500)
3. anti-DAZL, Rabbit, Polyclonal, Abcam, Cat. number ab34139, Lot number GR3184650-3 (1:200)
4. anti-5mC, Rabbit, Monoclonal, Cell Signaling Technology, Cat. number 28692, Lot number 2 (1:200)
5. anti-5mC, Mouse, Monoclonal, Abcam, Cat. number ab10805, Lot number GR3390032-1 (1:150)
6. anti-5hmC, Rabbit, Polyclonal, Active Motif, Cat. number 39769, Lot number 21518003 (1:500)
7. anti-DNMT3B, Sheep, Polyclonal, R&D Systems, Cat. number AF7646, Lot number CGWF0118031 (1:200)

8. anti-TFAP2C, Rabbit, Polyclonal, Santa Cruz Biotechnology, Cat. number sc-8977, Lot number H0715 (1:200)
9. anti-SOX9, Goat, Polyclonal, R&D Systems, Cat. number AF3075-SP, Lot number WIL0418111 (1:200)
10. anti-tdTomato, Goat, Polyclonal, SICGEN, Cat. number AB8181, Lot number 81221018 (1:100)
11. anti-DDX4, Rabbit, Monoclonal, Abcam, Cat. number 235442, Lot number GR3233447-1 (1:200)
12. anti-Mitochondria, Mouse, Monoclonal, Abcam, Cat. number ab92824, Lot number GR3307132-3 (1:800)
13. anti-SOX17, Goat, Polyclonal, R&D Systems, Cat. number AF1924, Lot number KGA1019061 (1:100)
14. APC conjugated SUSD2, Mouse, Monoclonal, BioLegend, Cat. number 327408, Lot number B218305 (1:100)
15. anti-TFCP2L1, Goat, Polyclonal, R&D Systems, Cat. number AF5726, Lot number CCUG0219121 (1:100)

The antibodies used for FACS:
1. Alexa Fluor 647-conjugated anti-alkaline phosphatase, Mouse, Monoclonal, BD Pharmingen, Cat. number 561500, Lot number 7132684 (5 µl/ sample)
2. Alexa Fluor 488-conjugated anti-alkaline phosphatase, Mouse, Monoclonal, BD Pharmingen, Cat. number 561495, Lot number 7132712 (5 µl/ sample)
3. PerCP-Cy5.5 conjugated anti-CDH5, Mouse, Monoclonal, BD Pharmingen, Cat. number 561566, Lot number B316764 (5 µl/ sample)
4. Alexa Fluor 647 conjugated anti-CD38, Mouse, Monoclonal, BioLegend, Cat. number 303514, Lot number B170786 (5 µl/ sample)

The antibodies used for WB:
1. anti-SOX17, Rabbit, Monoclonal, Cell Signaling Technology, Cat. number 81778, Lot number 1 (1:1000)
2. anti-PRDM1, Rabbit, Monoclonal, Cell Signaling Technology, Cat. number 9115, Lot number 6 (1:500)
3. anti-DMRT1, Rabbit, Monoclonal, Abcam, Cat. number ab126741, Lot number YI081707CS (1:1000)
4. anti-LaminB1, Rabbit, Polyclonal, Abcam , Cat. number ab16048, Lot number GR3244626-1 (1:1000)
5. horseradish-peroxidase-conjugated anti-rabbit IgG, Goat, Polyclonal, Agilent, Cat. number P0448, Lot number 20023997 (1:2000)

Validation

No home made antibodies were used in this study. All IF, WB and FACS antibodies were commercially validated as below:

anti-DMRT1, Abcam, Cat. number ab166893, Lot number GR119578-4, Clone number, EPR6936-209
Validation: https://www.abcam.com/products/primary-antibodies/dmrt1-antibody-epr6936-209-ab166893.html

anti-POU5F1, BD Biosciences, Cat. number 611203, Lot number 8087969, Clone number, 40/Oct-3
Validation: https://www.bdbiosciences.com/en-gb/products/reagents/microscopy-imaging-reagents/immunofluorescence-reagents/purified-mouse-anti-oct3-4.611203

anti-DAZL, Abcam, Cat. number ab34139, Lot number GR3184650-3
Validation: https://www.abcam.com/products/primary-antibodies/dazl-antibody-ab34139.html

anti-5mC, Cell Signaling Technology, Cat. number 28692, Lot number 2, Clone number, D3S2Z
Validation: https://www.cellsignal.com/products/primary-antibodies/5-methylcytosine-5-mc-d3s2z-rabbit-mab/28692

anti-5mC, Abcam, Cat. number ab10805, Lot number GR3390032-1, Clone number, 33D3
Validation: https://www.abcam.com/products/primary-antibodies/5-methylcytosine-5-mc-antibody-33d3-ab10805.html"

anti-5hmC, Active Motif, Cat. number 39769, Lot number 21518003
Validation: https://www.activemotif.com/catalog/details/39769/5-hydroxymethylcytidine-5-hmc-antibody

anti-DNMT3B, R&D Systems, Cat. number AF7646, Lot number CGWF0118031
Validation: https://www.rndsystems.com/products/human-dnmt3b-antibody_af7646

anti-TFAP2C, Santa Cruz Biotechnology, Cat. number sc-8977, Lot number H0715
Validation: https://www.scbt.com/p/ap-2gamma-antibody-h-77?productCanUrl=ap-2gamma-antibody-h-77&_requestid=1379409

anti-SOX9, R&D Systems, Cat. number AF3075-SP, Lot number WIL0418111
Validation: https://www.rndsystems.com/products/human-sox9-antibody_af3075

anti-tdTomato, SICGEN, Cat. number AB8181, Lot number 81221018
Validation: https://www.labome.com/product/SICGEN/AB8181-200.html

anti-DDX4, Abcam, Cat. number 235442, Lot number GR3233447-1, Clone number, EPR21789
Validation: https://www.abcam.com/products?keywords=235442

anti-Mitochondria, Abcam, Cat. number ab92824, Lot number GR3307132-3, Clone number, 113-1
Validation: https://www.abcam.com/products/primary-antibodies/mitochondria-antibody-113-1-bsa-and-azide-free-ab92824.html

anti-SOX17, R&D Systems, Cat. number AF1924, Lot number KGA1019061
Validation: https://www.rndsystems.com/products/human-sox17-antibody_af1924

APC conjugated SUSD2, BioLegend, Cat. number 327408, Lot number B218305, Clone number, W5C5
Validation: https://www.biolegend.com/en-gb/products/apc-anti-human-susd2-antibody-8410

anti-TFCP2L1, R&D Systems, Cat. number AF5726, Lot number CCUG0219121
Validation: https://www.rndsystems.com/products/human-tfcp2l1-antibody_af5726

Alexa Fluor 647-conjugated anti-alkaline phosphatase, BD Pharmingen, Cat. number 561500, Lot number 7132684, Clone number, B4-78
Validation: https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/alexa-fluor-647-mouse-anti-human-alkaline-phosphatase.561500

Alexa Fluor 488-conjugated anti-alkaline phosphatase, BD Pharmingen, Cat. number 561495, Lot number 7132712, Clone number, B4-78
Validation: https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/alexa-fluor-488-mouse-anti-human-alkaline-phosphatase.561495

PerCP-Cy5.5 conjugated anti-CDH5, BD Pharmingen, Cat. number 561566, Lot number B316764, Clone number, 55-7H1
Validation: https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/percp-cy-5-5-mouse-anti-human-cd144.561566

Alexa Fluor 647 conjugated anti-CD38, BioLegend, Cat. number 303514, Lot number B170786, Clone number, HIT2
Validation: https://www.biolegend.com/en-gb/products/alexa-fluor-647-anti-human-cd38-antibody-3264

anti-SOX17, Cell Signaling Technology, Cat. number 81778, Lot number 1, Clone number, D1T8M
Validation: https://www.cellsignal.com/products/primary-antibodies/sox17-d1t8m-rabbit-mab/81778

anti-PRDM1, Cell Signaling Technology, Cat. number 9115, Lot number 6, Clone number, C14A4
Validation: https://www.cellsignal.com/products/primary-antibodies/blimp-1-prdi-bf1-c14a4-rabbit-mab/9115

anti-DMRT1, Abcam, Cat. number ab126741, Lot number YI081707CS, Clone number, EPR6936
Validation: https://www.abcam.com/products/primary-antibodies/dmrt1-antibody-epr6936-ab126741.html

anti-LaminB1, Abcam , Cat. number ab16048, Lot number GR3244626-1
Validation: https://www.abcam.com/products/primary-antibodies/lamin-b1-antibody-nuclear-envelope-marker-ab16048.html

horseradish-peroxidase-conjugated anti-rabbit IgG, Agilent, Cat. number P0448, Lot number 20023997
Validation: https://www.agilent.com/en/product/specific-proteins/elisa-kits-accessories/goat-anti-rabbit-immunoglobulins-hrp-affinity-isolated-2717113

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	NANOS3-tdTomato and DMRT1-Venus reporter WIS2 line, NANOS3-Venus and DMRT1-tdTomato reporter WIS2 line, DAZL-tdTomato reporter WIS2 line, DMRT1-tdTomato reporter Shef-6 line, DAZL-tdTomato reporter Shef-6 line, NANOS3-tdTomato and DMRT1-Venus reporter WIS2 line bearing Dex-inducible SOX17/dox inducible PRDM1 transgenes, and DAZL-tdTomato reporter WIS2/Shef-6 lines bearing Dex-inducible SOX17/dox inducible DMRT1 transgenes were generated in this study according to the methods section. HEK 293 cells (CRL-1573) were obtained from ATCC.
Authentication	Authentication of ESCs and PGCLCs by PCR, qPCR, RNAseq and immunofluorescence.
Mycoplasma contamination	All lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used for this study.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Human embryonic genital ridges were collected from 2 individual embryos (wk7 Male and wk8 Female) for FACS analysis as shown in Figure 1b. Patient's identity is anonymised and authors have no access to donors' meta data.
Recruitment	Patients (who had already decided to undergo the termination of pregnancy operation) fully and freely consented to donate the foetal tissues for medical and academic research. Medical or surgical termination of pregnancy was carried out at Addenbrooke's Hospital, Cambridge, UK.
Ethics oversight	Human embryonic tissues were used under permission from National Health Service Research Ethical Committee, UK (REC Number: 96/085)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE223036 using the reviewer token - grgruemubzsxfwp

Files in database submission	Fastq files of three DMRT1 CUT&RUN libraries. We provided the peaks information (BED) in Table 3.
Genome browser session (e.g. UCSC)	Bigwig files for the replicates are available in GSE223036. IGV Genome browser can be used for visualisation.

Methodology

Replicates	Three (d4 DZ+PGCLC), Two (d8 DZ+PGCLC)
Sequencing depth	2 × 150 bp paired-end run. DMRT1 CUT&RUN in d4 DZ+PGCLC Replicate 1 : 31912103 total number of reads, 17557224 uniquely mapped reads Replicate 2 : 33388562 total number of reads, 19567846 uniquely mapped reads Replicate 3 : 13604948 total number of reads, 12065492 uniquely mapped reads DMRT1 CUT&RUN in d8 DZ+PGCLC Replicate 1 : 56903866 total number of reads, 48185963 uniquely mapped reads Replicate 2 : 51774414 total number of reads, 42507284 uniquely mapped reads Normal Rabbit IgG CUT&RUN in d8 DZ+PGCLC Replicate 1 : 43518396 total number of reads, 36337216 uniquely mapped reads
Antibodies	Rabbit monoclonal [EPR6936] to DMRT1 (Abcam, ab126741), Normal Rabbit IgG (#2729, Cell signaling)
Peak calling parameters	Peak calling parameters have been described in the method section.
Data quality	CUTandRUN libraries were assessed stringently by number of peaks and genome browser visualisation.
Software	All software information is described in the method section.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Aggregates were trypsinized with Trypsin/EDTA (0.25%, Thermo Fisher) at 37 degrees for 5-15 min and single cell suspension was incubated with conjugated antibodies for 30 min at room temperature.
Instrument	BD LSRFortessa Cell Analyzer (BD Bioscience)
Software	FlowJo
Cell population abundance	At least 10000 live single cell population was analyzed. The abundance of key populations is reported in each figure panels.
Gating strategy	Cell populations were gated first based on the SSC-A/FSC-A to exclude cell debris and dead cells, followed by gating single cell population based on FSC-A/FSC-W and live cell population based on DAPI

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.