A PAX5–OCT4–PRDM1 developmental switch specifies human primordial germ cells

Fang Fang1,2*, Benjamin Angulo1,2, Ninuo Xia1,2, Meena Sukhwani3, Zhengyuan Wang4, Charles C. Carey5, Aurélien Mazurie5, Jun Cui1,2, Royce Wilkinson5, Blake Wiedenheft5, Naoko Irie6, M. Azim Surani6, Kyle E. Orwig3 and Renee A. Reijo Pera1,2

Dysregulation of genetic pathways during human germ cell development leads to infertility. Here, we analysed bona fide human primordial germ cells (hPGCs) to probe the developmental genetics of human germ cell specification and differentiation. We examined the distribution of OCT4 occupancy in hPGCs relative to human embryonic stem cells (hESCs). We demonstrated that development, from pluripotent stem cells to germ cells, is driven by switching partners with OCT4 from SOX2 to PAX5 and PRDM1. Gain- and loss-of-function studies revealed that PAX5 encodes a critical regulator of hPGC development. Moreover, an epistasis analysis indicated that PAX5 acts upstream of OCT4 and PRDM1. The PAX5–OCT4–PRDM1 proteins form a core transcriptional network that activates germline and represses somatic programmes during human germ cell differentiation. These findings illustrate the power of combined genome editing, cell differentiation and engraftment for probing human developmental genetics that have historically been difficult to study.

Substantial research has centred on the identification and characterization of genes that are required for the specification, maintenance and differentiation of the mammalian PGCs that ultimately give rise to the sperm and eggs required to perpetuate life1–13. In mice, several transcription factors have been identified that are required for the in vitro specification and induction of the earliest stages of germ cells. These germ cells are ultimately able to mature and fulfill the greatest test of germ cell identity, the ability to produce live offspring14–18. However, the transcriptional network of hPGCs differs substantially from that of mice, making it difficult to translate this knowledge directly to humans19. For example, hPGCs express lineage specifier genes, such as SOX17, that are not expressed in mouse PGCs19.

Although hPGCs are committed to the germ cell lineage, they share with hESCs expression profiles of several pluripotency genes, including OCT4 (also known as POU5F1); however, other key pluripotency genes, such as SOX2, are not expressed in hPGCs11,20,21. How the network of pluripotency genes, which encode transcription factors, functions differently in hESCs and hPGCs is a fundamental question in the field of human germ cell developmental genetics that has remained unaddressed. Thus, in this study, we elucidated genetic mechanisms that underlie the development of hPGCs by identifying transcription factors that might function in a network to mediate hPGC specification and differentiation. We focused on OCT4, an essential gene that encodes a transcription factor that is expressed in both hESCs and hPGCs and is required to maintain cell identity in both cell types18,22–24. We developed methods to map the genome-wide binding of the OCT4 protein in highly heterogeneous tissue samples and identified OCT4 protein partners in hPGCs. We then used gain- and loss-of-function gene analyses to probe the function of PAX5, a member of the paired box (PAX) family, and discovered that it encodes a critical component of a genetic switch that is required for the transition from pluripotent stem cells to the differentiation of hPGCs.

Results

Global redistribution of OCT4 occupancy in the transition from hPGCs to hPGCs. To probe the role of OCT4 in hPGCs, we performed OCT4 chromatin immunoprecipitation (ChIP) assays with sequencing (ChIP–seq) on germ cells from second trimester human fetal testis cells. At this developmental stage, hPGCs have colonized the testis and are in the process of expanding to approximately 1–2 million cells in total, but have not differentiated to spermagonia or spermatocytes6. We noted that OCT4-positive cells are only present in the seminiferous tubules of the testis and not within the interstitial spaces (Fig. 1a, Supplementary Fig. 1a). Immunostaining data also indicated that OCT4-positive cells are a subpopulation of c-KIT-positive cells and do not express DDX4, which is an evolutionarily conserved germ cell marker of later stages of development (post-PGC; Supplementary Fig. 1b,c). However, since only 1% of the cells in the human fetal testis are OCT4-positive hPGCs (Fig. 1a, Supplementary Fig. 1a), and conventional ChIP protocols require a large number of homogenous cells, we adapted protocols from carrier ChIP65 and tissue ChIP66 to detect binding specificities of individual transcription factors within a heterogeneous cell mixture. We validated our protocol using a heterogeneous control mixture of 10,000 OCT4-positive hESCs mixed with 990,000 OCT4-negative fibroblast cells to model the composition of fetal testis (Supplementary Fig. 1d). We compared these data to that generated by conventional ChIP on a pure population of 1 million hESCs by quantitative PCR (qPCR; Supplementary Fig. 1e) and ChIP–seq. We found that the result from mixed-ChIP correlated highly with...
that from conventional ChIP (Supplementary Fig. 1f–h). Thus, our methods are reliable for generating binding data from a heterogeneous mixture of cells when coupled with highly specific antibodies.

We then applied the mixed-ChIP protocol to human fetal testis samples and generated a global binding profile for OCT4 in bona fide hPGCs. Two biological replicates were used and these replicates demonstrated a gene overlap of >90% (Supplementary Fig. 1h).

Although the enrichment profile of OCT4 around the transcription start sites was similar in both hPGCs and hESCs (Fig. 1b), there was a substantial redistribution of OCT4 binding. This redistribution was characterized by reduced binding near pluripotency-related genes (for example, OCT4, NANOG and LIN28A) and an enrichment of binding near germ cell-related genes (for example, PIWIL1, DDX4 and NANOS2) in hPGCs relative to hESCs (Fig. 1c,d). Furthermore,
to the germ cell lineage in an in vitro differentiation system. We demonstrated that the cell fate change from pluripotent hESCs to bona fide germ line hPGCs is associated with a global reorganization of OCT4 occupancy.

OCT4 switches partners to PAX5 and PRDM1 in hPGCs. We next set out to determine whether the genomic redistribution of OCT4 could be due to alternative OCT4 binding partners in hPGCs relative to hESCs. Immunostaining and RNA expression data showed that the expression of SOX2, OTX2 and ZIC2, genes that encode well-characterized functional partners of OCT4 in hESCs47-50, are significantly downregulated in hPGCs51 (Supplementary Fig. 2a,b). This result indicated that OCT4 might require different partners in hPGCs compared with in hESCs. To screen for potential OCT4-interacting transcription factors in hPGCs, we performed de novo sequence motif searches using OCT4-bound sequences exclusively in hPGCs and discovered motifs that are similar to consensus motifs for the transcription factors PAX5 and PRDM1 (Fig. 2a). Immunostaining demonstrated extensive co-expression of PAX5 and PRDM1 in OCT4-positive cells (Fig. 2b). Quantitation indicated that approximately 60% of the OCT4-positive cells are also positive for PAX5 and PRDM1. A RNA expression analysis also showed significant induction of PAX5 and PRDM1 in hPGCs compared with hESCs51 (Supplementary Fig. 2b). This result suggested that PAX5 and PRDM1 might co-occupy genomic loci with OCT4 as functional complexes in a germ cell-specific transcriptional network.

To determine the binding profiles of PAX5 and PRDM1 and to probe the potential association with OCT4 in hPGCs, we performed a ChiP–seq analysis for both PAX5 and PRDM1 (Supplementary Fig. 2c). Among the top 5,000 genes bound by each transcription factor, 1,441 genes, including germ cell-specific genes (for example, DDX4, DAZL and PIWIL1) were collectively bound by all three transcription factors (Fig. 2c,d). These results suggested extensive co-occupancy of PAX5, PRDM1 and OCT4 in hPGCs. An annotation analysis revealed that these co-bound genes are enriched for GO terms of germ cell signalling, hESC pluripotency and bone morphogenetic protein (BMP) signalling pathways (Fig. 2e). In addition, we observed that a recombinant OCT4–GST (glutathione S-transferase) fusion protein can immunoprecipitate recombinant PAX5 protein in vitro (Fig. 2f), indicating that PAX5 may interact directly with OCT4 protein in hPGCs. We did not detect direct protein interactions between recombinant OCT4 and PRDM1 proteins. However, PRDM1 protein was pulled down by PAX5 and vice versa (Supplementary Fig. 2d), indicating that OCT4, PAX5 and PRDM1 may assemble as a protein complex through both direct and indirect interactions. In summary, our observations suggested that OCT4, partnering with PAX5 and PRDM1 proteins, might constitute an extensive and unique transcription network in hPGCs.

Overexpression of PAX5 and PRDM1 induces human germ cell differentiation. Although PRDM1 is a well-studied key regulator of PGC development in both mice and humans52,53,54, the discovery of PRDM1 as a binding partner of OCT4 has not been previously shown. Additionally, although PAX5 is most commonly known for its role in development of the blood system55, a potential role for PAX5 in PGC development or specification has not yet been reported. To investigate the roles of both genes in hPGCs, we examined whether overexpression of PAX5 and PRDM1 is capable of directing hESCs to the germ cell lineage in an in vitro differentiation system. We overexpressed PAX5 or PRDM1 by approximately 10–15 times above the endogenous expression levels (Supplementary Fig. 3a) to a level comparable to that of bona fide hPGCs (Supplementary Fig. 2b). Forced expression of either PAX5 or PRDM1 did not alter the identity of ESCs in routine hESC maintenance (Supplementary Fig. 3b,c). Next, we differentiated hESCs overexpressing PAX5 or PRDM1 by exposing the cells to BMPs for 7 days, as previously described54. The expression of germ cell marker genes, including DAZL, DDX4, DPPA3 and NANOS3, was significantly upregulated in cells overexpressing these factors compared with control cells (Fig. 3a, Supplementary Fig. 3d). PAX5 also demonstrated binding on several later stage germ cell genes, such as SYCP1 and SYCP3, and activated their expression during differentiation (Fig. 3a, Supplementary Fig. 3e). Interestingly, co-expression of PAX5 and PRDM1 showed no obvious additive effects relative to overexpression of either gene alone (Supplementary Fig. 3f). Immunostaining revealed a DDX4 signal in cells overexpressing PAX5 but not control cells (Fig. 3b). We also used a hESC–DDX4–mOrange knock-in reporter cell line to better characterize the effects of in vitro differentiation. FACS analyses revealed a significantly higher percentage of DDX4–mOrange-positive cells in the cells overexpressing PAX5 relative to control cells (Fig. 3c). These data demonstrated that induced expression of PAX5 and PRDM1 in hESCs strongly promotes the differentiation of germ cells in vitro and prompted us to explore whether these cells may further mature if placed in the somatic niche via xenotransplantation.

To investigate the differentiation potential in vivo, we used a previously developed xenotransplantation platform56. Briefly, we transplanted green fluorescent protein (GFP)-tagged human cells into basulflan-treated immunodeficient mice, which are depleted of endogenous germ cells (Fig. 3d, Supplementary Fig. 4a). Two months post-transplantation, we analysed the testes and observed significant human germ cell engraftment in the tubules for cells overexpressing PAX5 or PRDM1, as indicated by the presence of GFP-positive cells that co-expressed DDX4 (Fig. 3e). More importantly, cells overexpressing PAX5 differentiated to a later stage germ cell fate and expressed mature germ cell markers, such as DAZL and DAZ1 (Fig. 3f). To quantify germ cell potential, we counted GFP-positive and DDX4-positive human germ cells and tubules across entire cross-sections. We then calculated the percentage of positive tubules and determined the number of GFP-positive and DDX4-positive cells in each positively stained tubule. Both values were significantly higher in cells overexpressing PAX5 or PRDM1 (Fig. 3g,h). Consistent with the in vitro differentiation results, cells overexpressing both PAX5 and PRDM1 promoted germ cell differentiation of hESCs in the mouse seminiferous tubule; however, no additive effects were detected compared with cells overexpressing each factor alone (Supplementary Fig. 4b–d). These data provide strong evidence that overexpression of PAX5 and PRDM1 is able to greatly promote the differentiation potential of hESCs towards the germ cell lineage in vitro and in vivo.

Knockout of PAX5 or PRDM1 reduces the germ cell potential of hESCs. We further examined the role of PAX5 in germ cell differentiation by performing loss-of-function studies. We generated a PAX5-knockout hESC line via CRISPR (clustered regularly interspaced short palindromic repeats)–Cas9-based genome editing (Supplementary Fig. 5a,b). Knockout of PAX5 was confirmed by immunofluorescence and western blot analyses upon induced neuronal differentiation55,56 (Supplementary Fig. 5c,d). Moreover, a significant reduction of germ cell gene expression was observed after in vitro differentiation (Fig. 4a) and DDX4 was not detected (Fig. 4b), further indicating that germ cell differentiation is severely reduced in vitro. We then examined whether germ cells could be differentiated and maintained from these cell lines via in vivo xenotransplantation. PAX5-knockout cells gave rise to significantly fewer positive tubules with human germ cells and fewer GFP-positive and DDX4-positive cells in the positive tubules; most of the
Fig. 2 | Co-occurrence of OCT4 with PAX5 and PRDM1 in PGCs. a, The position weight matrix of an enriched motif found in OCT4 ChIP-seq data from PGCs. The motif resembles the binding motifs for PRDM1 and PAX5. b, Cross-section of a human fetal testis (22 weeks). Upper panel: immunostained for PAX5 (red) and OCT4 (green), and DAPI stained for nuclei (blue). Lower panel: immunostained for PRDM1 (red) and OCT4 (green), and DAPI stained for nuclei (blue). Enlarged panels on the right represent the region enclosed within the white broken lines of the far left panel. White arrows indicate co-localization of PAX5 and OCT4 or PRDM1 and OCT4. Scale bars (original images), 100 µm; (expanded images) 50 µm. Immunostaining experiments were independently repeated a minimum of three times and similar results were obtained. c, Venn diagram of unique and shared genes bound by OCT4, PRDM1 and PAX5 in PGCs. The number of genes bound exclusively by each transcription factor or co-bound by multiple transcription factors are labelled. d, Genome browser representation of ChIP-seq tracks for OCT4 (yellow), PAX5 (blue) and PRDM1 (green) at the TBX3 and PIWIL1 loci. Regions that are bound collectively by OCT4, PAX5 and PRDM1 in PGCs are highlighted by pink shaded boxes. Y-axes represent ChIP-seq signals in units of SPMR (signals per million reads). ChIP-seq were independently repeated twice and similar results were obtained. e, GO analysis of co-bound genes. The analysis was performed twice and similar results were obtained. f, GST pull-down assay performed using OCT4 and PAX5 recombinant proteins. Pull-down was repeated three times and similar results were obtained. Unprocessed scans of western blots are shown in Supplementary Fig. 8.
Fig. 3 | Overexpression PAX5 and PRDM1 enhances the germ cell potential of ESCs. a, Heatmap of FPKM (fragments per kilobase of transcript per million mapped reads) values for genes associated with germline programming (top) and pluripotency (bottom). OE, overexpression. b, Differentiated cells from hESCs and cells overexpressing PAX5 immunostained for DDX4 and DAPI. Scale bars represent 50 μm. Immunostaining experiments were independently repeated a minimum of three times and similar results were obtained. c, Schematic of the experimental design for xenotransplantation. Transplantations were performed by independently injecting GFP-tagged human cells directly into seminiferous tubules of busulfan-treated mouse testes that were depleted of endogenous germ cells. Testis xenografts were analysed by immunohistochemistry 2 months after injection. d, Immunohistochemical analysis of testis xenografts derived from cells overexpressing PAX5 or PRDM1 and of control H1 hESCs. In all panels, broken white lines indicate the outer edges of spermatogonial tubules and enlarged views are shown on the right. White asterisks represent GFP-positive and DDX4-positive donor cells near the basement membrane. Scale bars (original images), 50 μm; (expanded images) 50 μm. Immunostaining experiments were independently repeated a minimum of three times and similar results were obtained. e, Percentage of tubules positive for GFP-positive and DDX4-positive cells were calculated across multiple cross-sections (relative to the total number of tubules). Data represent the mean ± s.d. of n = 4 independent replicates. P values were calculated by two-tailed Student’s t-test. Source data for g and h are provided in Supplementary Table 2.
DDX4-positive cells were mouse germ cells regenerared after treatment (Fig. 4c). There was a greater than threelfold reduction of positive tubules and a greater than fivefold reduction of GFP-positive and DDX4-positive cells in the positive tubules (Fig. 4d,e), indicating that genetic knockout of PAX5 greatly reduced germ cell differentiation from hESCs.

Previous studies have revealed that PRDM1-knockout hESCs fail to develop into PGCs in vitro\textsuperscript{6}. To determine whether these cells are also compromised in terms of germ cell differentiation in vivo, we used previously reported PRDM1-knockout hESCs\textsuperscript{6} to test their ability to differentiate into germ cells in murine xenotransplants. We observed that PRDM1-knockout hESCs were severely deficient in germ cell differentiation in vivo (Fig. 4f), with the majority of tubules devoid of any human germ cell engraftment. Only a small number of tubules were observed, with sparse GFP-positive and DDX4-positive cells. Counts of the engraftment revealed a greater than tenfold reduction in the formation of human germ cells in the mouse tubules (Fig. 4g,h), resulting in a more severe defect in germ cell potential compared with PAX5-knockout cells.

Epistasis of PAX5, OCT4 and PRDM1 in hPGCs. To explore the molecular mechanism of PAX5 function in hPGCs, we re-analysed our ChIP–seq data and observed that the enhancers of OCT4, which are bound by OCT4 itself in hESCs, are bound by PAX5 in hPGCs (Fig. 5a). Thus, we hypothesized that one role of PAX5 is to regulate and maintain OCT4 expression, as germ cell differentiation proceeds and requires expression of OCT4 at moderate levels\textsuperscript{39}. We first tested this hypothesis by examining OCT4 expression during in vitro differentiation. Following BMP-induced differentiation, OCT4 expression in cells overexpressing PAX5 was substantially elevated relative to differentiated control cells (Fig. 5b). However, due to the developmental limitation of current in vitro differentiation, genes essential for later stage germ cells (for example, DDX4 and DAZL), including PAX5, cannot be induced to the functional level\textsuperscript{31,38,39} (Supplementary Fig. 6a,b). Thus, PAX5-knockout cells only exhibited a minor decrease in OCT4 expression relative to control differentiated cells for all three protocols (Supplementary Fig. 6c). To overcome this limitation, we sorted hPGCs that were formed in vivo in mouse seminiferous tubules via GFP and c-KIT immunostaining and analysed the effects of PAX5 knockout in vivo (Fig. 5c,d). Note that the niche of mouse seminiferous tubules provide a superior differentiation environment for hESCs to develop to more mature hPGCs that express genes essential for later stage germ cells, including DDX4 (Fig. 4e,f). In these in vivo-derived hPGCs, there was a significant downregulation of OCT4 expression in cells formed by PAX5-knockout cells, and, as expected, a significant upregulation of OCT4 expression in cells formed by cells overexpressing PAX5 compared with cells formed by control hESCs (Fig. 5c).

To further determine whether PAX5 regulates OCT4 expression by regulating the enhancer of OCT4, a luciferase reporter assay was
performed in HEK293T cells. As expected, overexpression of PAX5 caused a significant increase in luciferase activity, suggesting that PAX5 could activate OCT4 expression through its enhancer (Fig. 5f). We also identified the binding motif of PAX5 in the region of the OCT4 enhancer sequences (Supplementary Fig. 7a–c). The results indicated that a mutation in the PAX5-binding motif abolished the induction effect of the PAX5 protein (Supplementary Fig. 7d).

Further analysis of the Chip-seq data indicated that PAX5 and OCT4 bind to PRDM1 enhancers with high intensity (Fig. 6a), suggesting that PAX5 and OCT4 might act upstream of PRDM1 to regulate its expression. We found an increase of PRDM1 in hPGCs formed from cells overexpressing PAX5 and a significant decrease in hPGCs formed from PAX5-knockout cells (Fig. 6c). The luciferase assay in HEK293T cells showed that both PAX5 and OCT4 were able to significantly increase luciferase activity driven by the PRDM1 enhancer (Fig. 6d). These results indicated that PAX5 and OCT4 could act on the PRDM1 enhancer as activators to induce PRDM1 expression, potentially during germine differentiation. A mutation in the PAX5-binding motif in the PRDM1 enhancer region abolished the induction effects of the PAX5 protein (Supplementary Fig. 7e–h). Since PRDM1 is a critical gene for germ cell specification and it could be downstream of PAX5 and OCT4, we overexpressed PRDM1 in PAX5-knockout cells to test whether PRDM1 could rescue the defect of PAX5-knockout cells. Indeed, we observed that overexpression of PRDM1 restored the germ cell potential of PAX5-knockout cells (Fig. 6e).

Taken together, our data shed light on the epistasis of these three transcription factors during differentiation from hESCs to hPGCs.
(Fig. 6f). In pluripotent stem cells, OCT4 interacts with SOX2 and other cofactors, and binds to its own enhancer to activate and maintain high expression. In contrast, during the differentiation of germ cells, PAX5 replaces OCT4, recognizes its own binding motif and binds to the enhancer of OCT4 to maintain a moderate expression of OCT4. Concurrently, PAX5 and OCT4 may bind to the enhancer

Fig. 6 | PAX5 and OCT4 act upstream of PRDM1. a, Genome browser representation of ChIP-seq tracks at the PRDM1 locus. Enhancer regions bound by OCT4 and PAX5 in PGCs are highlighted by pink shaded boxes. Y-axes represent ChIP-seq signals in units of SPMR. ChIP-seq was independently repeated twice and similar results were obtained. b, PRDM1 expression in control cells, cells overexpressing PAX5 and PAX5 knockout cells during in vitro differentiation. Data are represented as mean ± s.d. of n = 3 independent replicates. P values were calculated by two-tailed Student’s t-test. c, RT-qPCR analysis of PRDM1 expression in hPGCs formed in the mouse seminiferous tubules by PAX5 overexpressing, PAX5 knockout and control hESCs. Data are represented as mean ± s.d. of n = 3 independent replicates. P values were calculated by two-tailed Student’s t-test. d, Reporter construct used for measuring PRDM1 enhancer activity is shown. The genomic fragment bound by PAX5 and OCT4 (red) was inserted upstream of a luciferase gene driven by a minimal promoter. The y axis represents the fold enrichment of luciferase activity. Data represent the mean ± s.d. of n = 3 independent replicates. e, RT-qPCR analysis of the expression of genes associated with germline programming. Data represent the mean ± S.D. of n = 3 independent replicates. P values were calculated by two-tailed Student’s t-test. f, Model for gene regulation in pluripotency and germline programmes. In pluripotent stem cells, OCT4, together with other transcription factors (TFs) and cofactors, binds to its own enhancer to activate and maintain its high expression. While differentiating towards germine cells, PAX5 replaces OCT4 and binds to the enhancer of OCT4 to maintain a moderate expression of OCT4. Meanwhile, PAX5 and OCT4 bind to the enhancer of PRDM1 and activate its expression to initiate the germline programme. Source data for b, c, d and e are provided in Supplementary Table 2.
**Fig. 7 | Role of PAX5 and PRDM1 in hPGC specification in vitro.** a–c, RT-qPCR analyses of gene expression in all three germ layers. Comparison between H1 hESCs and cells overexpressing PAX5 (a), PAX5 knockout cells (b), and PRDM1 knockout cells (c) after BMP-induced differentiation. Data represent the mean ± s.d. of n = 3 independent replicates. P values were calculated by two-tailed Student’s t-test. d, Proposed molecular model for a transcriptional network centred by PAX5, OCT4 and PRDM1 in hPGCs. Upon induced germ cell differentiation with BMPs, OCT4 expression is reduced to moderate levels and maintained in partnership with PAX5. To efficiently induce germline programmes, OCT4 represses ectodermal genes and, at the same time, together with PAX5, activates PRDM1 to repress mesodermal and endodermal genes. In PAX5 knockout cells, OCT4 expression has decreased to levels so low that the expression of ectodermal genes is not suppressed effectively. Thus, the efficiency of induction of germ cells is low in PAX5 knockout cells and lower in PRDM1 knockout cells. Due to the low expression of OCT4 and loss of PRDM1 function, genes in all somatic lineages are upregulated and germ cell programmes fail to be activated. e, Summary of data establishing the roles of PAX5 and PRDM1 in hPGC specification in vitro and in vivo. The identity of hESCs is maintained by a core transcriptional network centred by OCT4, SOX2 and NANOG. Induced by BMP signals in vitro or in vivo by xenotransplantation, hESCs start to differentiate to early hPGCs, which express early germ cell markers, such as OCT4, SOX17, PRDM1 and NANOS3 (grey line with arrowhead). Overexpression of PAX5 is able to enhance the efficiency to early hPGCs and promote early hPGCs to the later stage, which express mature germ cell markers, such as DDX4, DAZL and DAZ1 (black line with arrowhead). Loss of PAX5 significantly reduces the germ cell potential of hESCs (grey broke line with arrowhead), while loss of PRDM1 leads to failure of hPGC specification (red line with an end bar). Source data for a–c are provided in Supplementary Table 2.
of PRDM1 and activate its expression to initiate the germ cell programme.

Molecular model of the PAX5–OCT4–PRDM1 network. OCT4 is known to repress ectoderm formation from hESCs and during germ cell differentiation (Fig. 1e). In addition, PRDM1 has been shown to suppress endoderm and other somatic genes during germ cell specification. Thus, we wondered whether PAX5, together with OCT4 and PRDM1, might function globally during germ cell differentiation. We examined the expression of somatic genes during in vitro differentiation (Supplementary Table 1). We detected downregulation of somatic genes belonging to the three primary germ layers in cells overexpressing PAX5 (Fig. 7a). In contrast, we observed significant upregulation of somatic genes in all three germ layers during the differentiation of PRDM1-knockout cells (Fig. 7c), confirming the role of this gene in suppressing the differentiation of somatic lineages during human germ cell development.

Based on our data, we propose a molecular model of human germ cell development (Fig. 7d). Upon external signalling via factors such as BMPs, germ cell differentiation is induced under both in vitro and in vivo conditions, and OCT4 expression is reduced to a moderate level that is partly maintained by a partnership with PAX5. To efficiently induce and maintain germline programmes, OCT4 represses ectodermal genes and at the same time, together with PAX5, activates PRDM1 to repress mesodermal and endodermal genes. In PAX5-knockout cells, OCT4 expression was low so that the expression of ectodermal genes was not suppressed efficiently. Thus, the efficiency of induction of germ cells is very low in PAX5-knockout cells. A more severe case is observed in PRDM1-knockout cells, in which there is an almost complete loss of expression of OCT4 and PRDM1. Moreover, genes in all somatic lineages are upregulated and the germ cell programmes fail to be activated.

Discussion

This work prompts a molecular model for germ cell development (Fig. 7e). In hESCs, OCT4 partners with pluripotent master regulators, including SOX2, to form the core transcriptional network that governs self-renewal and pluripotency. Induced by BMP signals in vitro and in vivo, hESCs differentiate into early hPGCs with low efficiency. Induction of PAX5 expression in hESCs results in their differentiation into more mature human germ cells, with significantly higher efficiency. These later stage differentiated cells closely resemble late hPGCs or gonocytes and express genes such as DDX4, KIT and DAZL, which mark mature human germ cells. Conversely, loss of function of PAX5 results in a lower efficiency in germ cell differentiation and loss of function of PRDM1 results in a complete failure of germ cell specification. Thus, this study provides evidence that human cell fate determination, at the juncture of pluripotency and somatic and germ line differentiation, may be the result of a balance of forces. That is, the co-expression of pluripotency genes (for example, OCT4 and NANOG) simultaneously with lineage specifiers (for example, SOX17, PAX5 and PRDM1) distinguishes hPGCs from all other human cells and from mouse PGCs. To maintain cell identity, hPGCs require a precise regulation and balance of pluripotency and lineage specifiers to move forward from pluripotent stem cell while repressing somatic lineage development and activating germ cell programmes. The PAX5–OCT4–PRDM1 axis that has been identified, along with the transcription factor genome-wide binding profiles, define the identity of bona fide hPGCs at stages beyond those commonly reported in vitro. The results of these studies may shed light on genetic requirements for human germ cell differentiation, enable more faithful and efficient production of human germ cells in vitro, and contribute to knowledge and models of human germ cell pathologies.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41556-018-0094-3.

Received: 5 January 2018; Accepted: 23 March 2018;
Published online: 30 April 2018

References

1. Ramathal, C. et al. DDX3Y gene rescue of a Y chromosome AZFa deletion rescues germ cell formation and transcriptional programs. Sci. Rep. 5, 15041 (2015).
2. Dominguez, A., Chiang, H., Sukhwani, M., Orwig, K. & Reijo Pera, R. A. Human germ cell formation in xenotransplants of induced pluripotent stem cells carrying X chromosome aneuploidies. Sci. Rep. 4, 6432 (2014).
3. Ramathal, C. et al. Fate of iPSCs derived from azoospermic and fertile men following xenotransplantation to seminiferous tubules. Cell Rep. 7, 1284–1297 (2014).
4. Durruthy, J. D. et al. Fate of induced pluripotent stem cells following transplantation to murine seminiferous tubules. Hum. Mol. Genet. 23, 3071–3084 (2014).
5. Wu, K., Angeles, V., Flores, M., Nguyen, H. & Reijo Pera, R. A. Human DAZL, DAZ and ROULE genes modulate primordial germ cell and haploid ganete formation. Nature 462, 222–225 (2009).
6. Salto Mamsen, L., Lutterodt, M. C., Andersen, E. W., Byakov, A. G. & Andersen, C. Y. Germ cell numbers in human embryonic and fetal gonads during the first two trimesters of pregnancy: analysis of six published studies. Hum. Reprod. 26, 2140–2145 (2011).
7. Clark, A. T. et al. Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. Hum. Mol. Genet. 13, 727–739 (2004).
8. Reijo Pera, R. A., Alagappan, R. K., Patrizio, P. & Page, D. C. Severe oligosperma resulting from deletions of the azoospermia factor gene on the Y chromosome. Lancet 347, 1290–1293 (1996).
9. Irie, N., Tang, W. W. & Azim Surani, M. Germ cell specification and pluripotency in mammals: a perspective from early embryogenesis. Reprod. Med. Biol. 13, 203–215 (2014).
10. Magnudottir, E. et al. A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat. Cell Biol. 15, 905–915 (2013).
11. Tang, W. W., 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).
12. Saitou, M., Barton, S. C. & Surani, M. A. A molecular programme for the specification of germ cell fate in mice. Nature 418, 293–300 (2002).
13. Saitou, M. & Yamaji, M. Primordial germ cells in mice. Cold Spring Harb. Perspect. Biol. 4, a008375 (2012).
14. Nakaki, F. et al. Induction of mouse germ-cell fate by transcription factors in vitro. Nature 501, 222–226 (2013).
15. Ohnata, Y. et al. Bmp1 is a critical determinant of the germ cell lineage in mice. Nature 436, 207–213 (2005).
16. Weber, S. et al. Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biol. Reprod. 82, 214–223 (2010).
17. Yamaji, M. et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat. Genet. 40, 1016–1022 (2008).
18. Zhou, Q. et al. Complete meiosis from embryonic stem cell-derived germ cells in vitro. Cell Stem Cell 18, 330–340 (2016).
19. Irie, N. et al. SOX17 is a critical specifier of human primordial germ cell fate. Cell 160, 253–268 (2015).
20. Tang, W. W. et al. A unique gene regulatory network resets the human germline epigenome for development. Cell 161, 1453–1467 (2015).
21. 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 (2017).
22. Kehler, J. et al. Oct4 is required for primordial germ cell survival. EMBO Rep. 5, 1078–1083 (2004).
23. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 (c-Kit) in embryonic stem cells in vitro. Nature 363, 2185–2195 (1990).
24. Zhou, Q. et al. Complete meiosis from embryonic stem cell-derived germ cells in vitro. Cell Stem Cell 18, 330–340 (2016).
25. Scholer, H. R., Dressler, G. R., Balling, R., Rohdehwold, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO J. 9, 2185–2192 (1990).
26. O’Neill, L. P., VerMilyea, M. D. & Turner, B. M. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. Nat. Genet. 38, 835–841 (2006).
26. Cotney, J. L. & Noonan, J. P. Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing. *Cold Spring Harb. Protoc.* 2015, 191–199 (2015).
27. Nishimoto, M., Fukushima, A., Okuda, A. & Muramatsu, M. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol. Cell. Biol.* 19, 5453–5465 (1999).
28. Boyer, L. A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956 (2005).
29. Buecker, C. et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14, 838–853 (2014).
30. Pardo, M. et al. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* 6, 382–395 (2010).
31. Gkountela, S. et al. The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat. Cell Biol.* 15, 113–122 (2012).
32. Vincent, S. D. et al. The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* 132, 1313–1325 (2005).
33. Medvedovic, J., Ebert, A., Tagoh, H. & Busslinger, M. Pax5: a master regulator of B cell development and leukemogenesis. *Adv. Immunol.* 111, 179–206 (2011).
34. Kee, K. & Reijo Pera, R. A. Human germ cell lineage differentiation from embryonic stem cells. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5048 (2008).
35. Zhang, P., Xia, N. & Reijo Pera, R. A. Directed dopaminergic neuron differentiation from human pluripotent stem cells. *J. Vis. Exp.* https://doi.org/10.3791/51737 (2014).
36. Perrier, A. L. et al. Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* 101, 12543–12548 (2004).
37. Yeom, Y. J. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 122, 881–894 (1996).
38. Sasaki, K. et al. Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* 17, 178–194 (2015).
39. Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* 10, 440–454 (2012).

**Acknowledgements**

This work was supported by P50 HD 068158 to R.A.R.P. (Project I).

**Author contributions**

The study was conceived and designed by FF and R.A.R.P. FF performed most experiments (including ChIP–seq, immunohistochemistry, RNA-seq, protein pull-down assays, luciferase reporter assays, gene expression profiling) and analysed the data. NX performed bioinformatics analyses for ChIP–seq and luciferase reporter assays, flow cytometry and gene expression analysis for the xenotransplantation experiments. BA generated the PAX5-knockout hESC lines and performed part of the immunohistochemistry in xenotransplantation samples. ZW performed the initial bioinformatics analysis for ChIP-seq data. MS and KE.O. conducted the xenotransplantation. CCC and AM performed RNA-seq analysis. JC constructed the H1-DDX4 reporter. RW and BW designed and constructed the PAX5-knockout plasmids. MAS and NI provided the PRDM1-knockout hESC line and protocol. The manuscript was written by FF and R.A.R.P. with input from the other authors.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-018-0094-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to FF.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

hESC culture and differentiation. hESC (WiCell) were maintained feeder-free on Matrigel (BD Biosciences)-coated plates as previously described\(^4\). The cell line was authenticated by immunostaining for cell-type-specific proteins and tested as negative for mycoplasma contamination. The cells were grown at 37°C with 5% CO\(_2\) in mTeSR1 medium (Stemcell Technologies). For differentiation, cells were seeded overnight at a density of 200,000 per well in 6-well plates and medium was replaced the next day with differentiation media (knockout DMEM supplemented with 20% fetal bovine serum, 1 mM L-glutamine, 0.1 mM nonessential amino acids, 0.1 mM β-mercaptoethanol and 50 mg/ml recombinant human BMP4, BMP7 and BMP8b (R&D systems)) as previously described\(^5,6\). Differentiation medium was changed every other day for 7 days.

 Xenotransplantation assay. Human cell lines were transplanted into the testes of busulfan-treated, immunodeficient nude mice (NCr nu/nu; Taconic) as described previously\(^7\). Mouse testes were replaced the next day with differentiation media (knockout DMEM supplemented with 0.1 mM L-glutamine and 5% CO\(_2\) in mTeSR1 medium (Stemcell Technologies)). The cell line was authenticated by immunostaining for cell-type specific proteins and tested as negative for mycoplasma contamination. For differentiation, cells were seeded overnight at a density of 200,000 per well in 6-well plates and medium was changed every other day for 7 days.

 Human cell lines were transplanted into the testes of busulfan-treated, immunodeficient nude mice (NCr nu/nu; Taconic) as previously described\(^7\). Mouse testes were replaced the next day with differentiation media (knockout DMEM supplemented with 0.1 mM L-glutamine and 5% CO\(_2\) in mTeSR1 medium (Stemcell Technologies)). The cell line was authenticated by immunostaining for cell-type specific proteins and tested as negative for mycoplasma contamination. For differentiation, cells were seeded overnight at a density of 200,000 per well in 6-well plates and medium was changed every other day for 7 days.

 For nuclear staining, 4,6-diamidino-2-phenylindole (DAPI; 1 μg/ml) was used. Images shown are representative of at least three independent experiments.

 RNA isolation, library preparation and sequencing analysis. Total RNA was extracted using an Acuturus PicoPure RNA Isolation kit (Life Technologies). RNA quality was determined using a Bioanalyzer 2100 (Agilent). Sequencing libraries were constructed using a SMARTer universal low input RNA kit (Clontech) according to the manufacturer’s instructions. DNA library samples were submitted to the Stanford Genomics Facility and 100-base pair-end high-throughput sequencing was performed. All sequenced libraries were mapped to the human genome using TopHat and Cufflink\(^8,9\) and the default parameter setup. Differential expression was analysed using StrandNGS (AvadisNGS).

 Gene expression analysis by qPCR. Total RNA for qPCR was extracted from cells using a RNeasy Plus Mini kit (Qiagen) and 1 μg RNA was reverse transcribed using the SuperScript III First-Strand Synthesis System. qPCR was performed in triplicate using Power SYBR Green PCR Master Mix (both from Life Technologies) with the data normalized to housekeeping genes. The primer sequences are listed in Supplementary Table 1. Data shown are representative of at least three independent experiments.

 Immunohistochemistry of recipient mouse testes. Paraformaldehyde-fixed mouse testes were sectioned by AML Laboratories with paraffin embedding and conducting serial cross-sectioning every 5 mm. For deparaffinization, two consecutive xylene (Sigma-Aldrich) treatments (5 min each) were performed followed by rehydration in 100%, 90%, 80% and 70% ethanol treatments. A 10-min wash in tap water was then performed. Antigen retrieval was conducted by boiling slides for 30 min in 0.01 M sodium citrate (pH 6.0; Sigma-Aldrich), cooling slides for 30 min, followed by a 10 min wash in PBS. Blocking and permeabilization was conducted by adding 10% normal goat serum (Vector ImmunoResearch) with 0.1% Triton X-100 (Sigma-Aldrich) in PBS for 1 h, followed by incubation with primary antibodies diluted in 1% blocking solution overnight at 4°C in a humidified chamber. The following antibodies were used: DDX4 (R&D; AF2030), Gfp (Abcam; ab13970), OCT4 (Santa Cruz; sc-8628) and DAZL (Novus; NB100-2437). Slides were washed in PBST, followed by incubation for 1 h with fluorescently labelled secondary antibodies raised in donkey, followed by additional washes in PBST. The primary antibodies were used at 1:20 dilution and secondary antibodies were used at 1:300 dilution. All samples were mounted using ProLong Gold Anti-fade mounting media containing DAPI (Life Technologies). Samples were then imaged using a confocal microscope (Zeiss). Quantification of GFP-positive and DDX4-positive staining was determined manually from multiple sections taken from two to three different depths within the testis (technical replicates) and from multiple clonal replicates for each transplanted cell line (biological replicates).

 Motif analysis. Motif analysis was performed using the MEME-Chip suite and the default parameter settings. The motif definition and motif selection was based on the motif ECH1 from the motif database.

 CRISPR design and PAX5-knockout derivation. CRISPR guide RNAs (gRNAs) were designed using the online CRISPR design tool from the Massachusetts Institute of Technology (http://crispr.mit.edu/). Candidate gRNAs with the highest score were chosen for each genomic region. Oligonucleotides for these gRNAs were synthesized and cloned into plasmid pX459 (Addgene; 48139) carrying both Cas9 and gRNA expression cassettes, with one modification of the original plasmid in which the Cas9 promoter was replaced by the EEF1A1 promoter. The cutting efficiency of each gRNA construct was validated by transfecting HEK293T cells and sequencing the target regions in the genome. CRISPR pairs were nucleoected into hESC cells and plated as single cells. Single cells were clonally expanded and isolated for PCR to test for successful PAX5 deletion and for sequencing.

 Construction of overexpression constructs by lentiviral vectors. The sequences of PAX5 and PRDM1 were assembled using Gibson Assembly Cloning technology (NEB). Briefly, gBlocks (gene fragments) were synthesized by Integrated DNA Technologies, and individual gBlocks were assembled to one gene transcript with Gibson Assembly Cloning technology followed by an amplification reaction with Phusion DNA polymerase according to the manufacturer’s instructions. Amplified gene fragments were ligated into a pENTR/D-TOPO vector (Life Technologies) using the Multisite Gateway system. Clones were transformed into One-Shot Competent Escherichia coli, and DNA was purified and sequenced. Positive clones were used for a recombination reaction with the Gateway destination vector (pcDNA-DEST40). Subsequent transformation into One-Shot Competent E. coli, followed by DNA purification and sequencing for verification of correct cloning, resulted in overexpression vectors for PAX5 and PRDM1.

 Luciferase assay for enhancer activity. Enhancer sequences were generated by PCR of human genomic DNA discovered for POU5F1 and PRDM1, and then cloned into pGL2.28(luc2CP/minP/Hygro) (Promega; E8461) with restricted cloning into pGL2.28(luc2CP/minP/Hygro) (Promega; E8461) with restricted cloning into pGL2.28(luc2CP/minP/Hygro) (Promega; E8461) with restricted
enzymes KpnI and BstI (NEB). The minimal promoter pGL4.28 was used as negative control. Luciferase activity was measured using a dual-luciferase reporter assay system (Promega) as per the manufacturer's protocol.

**GO analysis.** GO enrichment was performed using GREAT analysis and the default parameters46.

**FACS and flow cytometry.** Cells were dissociated in 0.25% trypsin–EDTA buffer (Gibco) at 37 °C for 5 min and collected by centrifugation at 200 g in an Eppendorf 5702R centrifuge. Then the cells were passed through 70 µm strainers (BD Biosciences) to ensure they were digested as single cells before they were subject to flow cytometry analysis. Mouse testes cells were dissociated with 0.25% trypsin–EDTA buffer for 30 min at 37 °C. Dissociated cells were incubated in 1% BSA in PBS containing primary antibodies (c-KIT (DAKO; A4502)) on ice for 20 min. Cells were then analysed for mOrange expression or c-KIT and GFP using a BD FACS Aria II cell sorter. Analysis was performed using LSRII (Beckton Dickinson) and FlowJo software (Tree Star).

**GST pull-down assay.** The recombinant protein OCT4 (Novus; H00005079-P01) was bound to glutathione–sepharose beads (Amersham) and incubated with recombinant PAX5 protein (Novus; H00005079-P01) overnight at 4 °C. Beads were washed six times with cell lysis buffer. The eluents were analysed by western blotting. The blots shown are representative of at least three independent experiments.

**Differentiation to neuronal progenitor cells from hESCs.** Differentiation of hESCs into neuronal progenitor cells was carried out as previously described35,36. hESCs were dissociated into single cells with accutase, depleted of mouse embryonic fibroblasts (MEF) feeders by incubating on a gelatin-treated culture dish for 30 min, and then plated onto a Matrigel-coated dish at a density of 36,000 cells cm–2 in mTeSR1 (Stemcell Technologies) in the presence of 2 µg ml–1 M thiazovivin (Santa Cruz Biotechnology). Differentiation was started 48 h later with KnockOut Serum Replacemnt (KSR) differentiation medium supplemented with different combinations of small molecules5. Each well of the 6-well plate was coated by 1 ml fibronectin, which was diluted with cold (4 °C) PBS to 2 µg ml–1, and the plates were incubated at 37 °C overnight before seeding the cells. After 20 days of differentiation, cells were replated into plates coated with poly-L-ornithine, laminin and fibronectin and cultured with B27 differentiation medium.

**Statistics and reproducibility.** No statistical methods were used to predetermine samples or outcomes. For the xenotransplantation studies, animals were randomly allocated into groups receiving various cell line injections. All statistical analyses were conducted using GraphPad Prism (version 5). Two-tailed Student’s t-test were used when data met criteria for parametric analysis (normal distribution or equal variances). P values are shown in the figures, and the number of biological replicates for each experiment is indicated in the figure legends. Experiments were repeated independently at least three times.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** All sequencing data that support the findings of this study have been deposited in NCBI’s GEO under accession code GSE100639. Previously published sequencing data that were re-analysed here are available under accession codes GSE60138, GSE39821 and GSE67259. Source data for the following figures are provided in Supplementary Table 2: Fig. 3g,h; Fig. 4a,d,e,g,h; Fig. 5d,e,g,h; Fig. 6a–c; Supplementary Fig. 2b; Supplementary Fig. 3a,d,f; Supplementary Fig. 4c,d; Supplementary Fig. 6a,b; and Supplementary Fig. 7a,c,d,e,g,h. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

**References.**
40. Kee, K., Gonsalves, J. M., Clark, A. T. & Reijo Pera, R. A. Bone morphogenetic proteins induce germ cell differentiation from human embryonic stem cells. Stem Cells Dev. 15, 831–837 (2006).
41. Hermann, B. P. et al. Spermatogonial stem cell transplantation into rhesus testes regenerates spermatogenesis producing functional sperm. Cell Stem Cell 11, 715–726 (2012).
42. Trappnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111 (2009).
43. Trappnell, C. et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515 (2010).
44. Sandelin, A. et al. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics 5, 99 (2004).
45. Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res. 24, 238–241 (1996).
46. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501 (2010).
Corresponding author(s): Fang Fang

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of 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
- A full description of the statistics 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
- Clearly defined error bars
- State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

- no software was used.

Data analysis

- For ChIP-seq analysis: Bowtie (1.1.2), MACS (1.4.1).
- For RNA-seq analysis: TopHat (2.1.1), Cufflink (2.2.1), StrandNGS.
- For Motif analysis: MEME-ChIP(4.12.0).
- Statistical analysis: GraphPad Prism 5 (5.03).
- Flow cytometry analysis: Flowjo (10.4.2)

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data
All sequencing data that support the findings of this study have been deposited in NCBI’s GEO under accession number GSE100639. Source data for Fig. 3g, h; 4a, d, e, g, h; 5d, e, g, h; and Supplementary Fig. 2b; 3a, d, f; 4c, d; 6a, b; 7a, c, d, e, g, h is provided in Supplementary Table 2. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Field-specific reporting

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

- Life sciences
- Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

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

| Sample size | No statistical was used to predetermine sample size. The sample size was determined based on the previous publications and experimental designs that have similar objectives (Ramathal et al., Cell Rep., 2014; Ramathal et al., Sci. Rep., 2015; Durruthy et al., Hum. Mol. Genet., 2014; Kee et al., Nature, 2009). Three to more independent results were used to perform statistical analyses. If less, no statistics were performed from these samples. All source data required for statistical tests are indicated in Supplementary Table 2 (Statistics data source). |
| Data exclusions | Data were excluded when negative or positive controls were not working. |
| Replication | For each experiment, we performed at least two independent biological replicates and all attempts were successful. |
| Randomization | For the xenotransplantation studies, animals were randomly allocated into groups receiving various cell line injections. Randomization (formal or otherwise) was not relevant for other data included in the manuscript. |
| Blinding | Investigators were not blinded to group allocation. For the counting analysis for xenotransplantation, investigators were blinded to the samples. |

Materials & experimental systems

Policy information about availability of materials

Involved in the study

- n/a
- Unique materials
- Antibodies
- Eukaryotic cell lines
- Research animals
- Human research participants

Antibodies

| Antibodies used |
|-----------------|
| anti-OCT4 (sc-8628, Santa Cruz Biotechnology, Lot#: C2613); |
| anti-PAX5 (sc-1974, Santa Cruz Biotechnology, Lot#: A3113); |
| anti-PRDM1 (C14A4, Cell Signaling Technology, Lot: 4); |
| anti-DDX4 (R&D #AF2030, Lot: KPX0 4051); |
| anti-GFP (Abcam; ab18373, Lot: GR254056-13); |
| anti-DAZL (Novus; NB100-2437; Lot: P11); |
| anti-KIT (A4502; DAKO); |
| anti-beta-ACTIN (8H10D10, Cell Signaling Technology, Lot: 15); |
| HRP-conjugated anti-rabbit IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2004, Lot: HD913; |
| HRP-conjugated anti-goat IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2020, Lot: B0613; |
| HRP-conjugated anti-mouse IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2005. |
Alexa Fluor® 488 AffiniPure Donkey Anti-Goat IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 705-545-003
Alexa Fluor® 594 AffiniPure Donkey Anti-Rabbit IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 711-585-152.
The commercial antibodies were validated based on the information on the manufacturers’ instructions.
For ChIP and Co-IP assay, antibodies were used at 1:100.
For western blot, primary antibodies were used at 1:1000 and secondary antibodies were used at 1:5000.
For immunostaining, primary antibodies were used at 1:200 and secondary antibodies were used at 1:300.
anti-OCT4 (sc-8628, Santa Cruz Biotechnology, Lot#: C2613);
anti-PAX5 (sc-1974, Santa Cruz Biotechnology, Lot#: A3113);
anti-DDX4 (R&D #AF2030, Lot: KPX0 4051);
anti-GFP (Abcam; ab18373, Lot: GR2S4056-13);
anti-DAZL (Novus; NB100-2437; Lot: P1);
anti-CKIT (A4502; DAKO);
anti-beta-ACTIN (BH10010, Cell Signaling Technology, Lot:15)
HRP-conjugated anti-rabbit IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2004, Lot: H0913.
HRP-conjugated anti-goat IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2020, Lot: B0613.
HRP-conjugated anti-mouse IgG secondary antibody, Supplier: Santa Cruz Biotechnology, Cat.: sc-2005.
Alexa Fluor® 488 AffiniPure Donkey Anti-Goat IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 705-545-003
Alexa Fluor® 594 AffiniPure Donkey Anti-Rabbit IgG (H+L), Jackson ImmunoResearch Laboratories, Code: 711-585-152.
For ChIP and Co-IP assay, antibodies were used at 1:100.
For western blot, primary antibodies were used at 1:1000 and secondary antibodies were used at 1:5000.

Validation

The commercial antibodies were validated based on the information on the manufacturers’ instructions.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) H1 hESCs (WiCell) (feeder free)
293T (ATCC)

Authentication

Immunostaining of cell type specific markers was performed for cell line authentication.

Mycoplasma contamination

All lines were tested negative for mycoplasma contamination.

Commonly misidentified lines

(See ICLAC register)

No commonly misidentified cell lines were used

Research animals

Policy information about studies involving animals. ARRIVE guidelines recommended for reporting animal research

Animals/animal-derived materials In this study, we used busulfan-treated, immune-deficient nude male mice (Ncr nu/nu; Taconic). Mice are treated with busulfan (40 mg/kg) at 5-6 weeks of age and then transplanted 5-12 weeks post busulfan.

Human research participants

Policy information about studies involving human research participants

Population characteristics In this study, we obtained second trimester human fetal testis tissue. Note that the protocol for tissue procurement and use was approved by the Institutional Review Board of Montana State University (RR-P031014-EX); all procedures were compliant with all relevant ethical regulations regarding human research with unidentifiable banked tissue.

Method-specific reporting

n/a Involved in the study

ChIP-seq

Flow cytometry

Magnetic resonance imaging

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/submission/update/?acc=GSE100639

Files in database submission

- 1M_H9_OCT4_ChIP_CGATGT-04-30-2013.fastq
- 2nd_OCT4_SC_GCCAAT-04-30-2013.fastq
- 2nd_Test_OCT4_Ab_ACAGTG-04-30-2013.fastq
- IM_Mix_OCT4_TGACCA-04-30-2013.fastq
- testes_prdm1_chip-11-26-2013.fastq
- PAX5-testis-130909_BRISCOE_0116_BD2CRUACXX_L6_CTTGTA_pf.fastq.gz
- input-11-26-2013.fastq
- 1M_H9_OCT4.sorted.dedup.bam
- 2nd_OCT4_SC.sorted.dedup.bam
- 2nd_Test_OCT4_Ab.sorted.dedup.bam
- IM_Mix_OCT4.sorted.dedup.bam
- PAX5_TESTS_BRISCOE_0116_BD2CRUACXX_L6_CTTGTA_pf.bam
- TESTES_PRDM1.bowtie2.srt.bam

Genome browser session
(e.g. UCSC)

no longer applicable

Methodology

Replicates
We did two replicates for OCT4 ChIP in human fetal testis, for the other ChIP, we only provided one

Sequencing depth
single-end, 36bp

Antibodies
anti-OCT4 (sc-8628, Santa Cruz Biotechnology);
anti-PAX5 (sc-1974, Santa Cruz Biotechnology);
anti-PRDM1 (C14A4, Cell Signaling Technology);

Peak calling parameters
Effective genome size 2700000000.0
Tag size 49
Band width 300
Pvalue cutoff for peak detection 1e-05
Save shifted raw tag count at every bp into a wiggle file wig
Extend tag from its middle point to a wigextend size fragment. -1
Resolution for saving wiggle files 10
Use fixed background lambda as local lambda for every peak region False
3 levels of regions around the peak region to calculate the maximum lambda as local lambda 1000,5000,10000
Build Model create_model
Diagnosis report no_diag
Perform the new peak detection method (futurefdr) False

Data quality
OCT4 ChIP in human fetal testis: 27061
PAX5 ChIP in human fetal testis: 22728
PRDM1 ChIP in human fetal testis: 48399
OCT4 ChIP in hESCs:14799

Software
Alignment by Bowtie-Illumina/Galaxy
Peak call by MACS-ChIP-seq/Galaxy
BiWig generated by WigToBigWig/Galaxy
Motif detection by MEMEChip

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
hESCs were dissociated in 0.25% trypsin–EDTA (Gibco BRL) at 37°C for 5 min and collected by centrifugation at 200g in an Eppendorf 5702 R centrifuge. Then the cells were passed through the 70µM strainers (BD Biosciences) to make sure they were
digested as single cells before they were subject to the flow cytometry. Mouse testis cells were dissociated with 0.25% Trypsin–EDTA for 5 min at 37°C. Dissociated cells were passed through the 40uM strainers (BD Biosciences) and then incubated in 1% BSA in PBS containing primary antibodies on ice for 20 min.

| Instrument     | BD FACS Aria 2.0 |
|----------------|------------------|
| Software       | For data collection: BD FACSDiva v 6.1.2 software; For data analysis: Flowjo (10.4.2) |
| Cell population abundance | Sorted cells were directly placed into DNA extraction buffer. The abundance for these samples could not be assessed. |
| Gating strategy | After cells were selected in the FSC/SSC dot plot to remove debris, they were gated to exclude cellular aggregates in the FSC/FSC dot plot. Gates of GFP-FITC, mOrange-PE, or CKIT-PE cells were set and compared with a control sample with no detectable fluorochrome expression. |

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