Monolayer platform to generate and purify human primordial germ cells in vitro provides new insights into germline specification

Vittorio Sebastiano (✉ vsebast@stanford.edu)
Stanford Medicine

Gugene Kang
Stanford University

Sivakamasundari Vijayakumar
Stanford University

Roberta Sala
Stanford University

Angela Chen
Stanford University

Abidemi Adebayo
Stanford University

Andrea Cipriano
Stanford University

Jonas Fowler
Stanford University

Lay Teng Ang
Stanford University

Kyle Loh
Stanford University School of Medicine   https://orcid.org/0000-0002-8042-0149

Abstract

Generating primordial germ cells (PGCs) from human pluripotent stem cells (hPSCs) advances studies of human reproduction and development of infertility treatments, but currently entails complex 3D aggregates. Here we develop a simplified, monolayer method to differentiate hPSCs into PGCs within 3.5 days. We used our simplified differentiation platform and single-cell RNA-sequencing to uncover new insights into PGC specification. Transient WNT activation for 12 hours followed by WNT inhibition specified PGCs; by contrast, sustained WNT instead induced primitive streak. Thus, somatic (primitive streak) and PGCs are related—yet distinct—lineages segregated by temporally-dynamic signaling. Pluripotency factors including NANOG are continuously expressed during the transition from pluripotency to posterior epiblast to PGCs, thus bridging pluripotent and germline states. Finally, hPSC-derived PGCs can be easily purified by virtue of their CXCR4$^+$PDGFRA$^-$GARP$^-$ surface-marker profile and single-cell RNA-sequencing reveals that they harbor strong transcriptional similarities with fetal PGCs.

Introduction

Within the mammalian embryo, Primordial Germ Cells (PGCs) are the harbinger to eggs and sperm; consequently, they are key to the act of reproduction and the vertical transmission of genetic and epigenetic information to the next generation. The developmental origins of mouse PGCs have been thoroughly explored (Chiquoine, 1954; Ginsburg et al., 1990; Lawson and Hage, 1994), thus enabling the stepwise differentiation of mouse pluripotent cells into PGC-like cells (PGCLCs) *in vitro* in 3D cultures (Hayashi et al., 2011; Ohinata et al., 2009). Pluripotent stem cells (PSCs) from both humans and non-human primates have likewise been successfully differentiated into PGCLCs in 3D cultures (Chen et al., 2017; Chen et al., 2019; Irie et al., 2015; Kobayashi et al., 2017; Kojima et al., 2017; Mitsunaga et al., 2017; Sakai et al., 2020; Sasaki et al., 2015; Sosa et al., 2018; Sybirna et al., 2020; Yokobayashi et al., 2017). The ability to generate human PGCLCs *in vitro* has shed light on early human germline specification as well as genetic diseases such as infertility (Fang et al., 2020; Zhao et al., 2018) and could eventually enable *in vitro* manufacture of human eggs and sperm for infertility treatments and other reproductive technologies.

Fundamentally speaking, the developmental precursors to human PGCs and the signals that specify their formation *in vivo* have remained hitherto uncertain because PGCs arise in weeks 2–3 of human embryogenesis (De Felici, 2016; O'Rahilly and Müller, 1987) and it is ethically and technically difficult to attain and analyze early post-implantation human embryos. Consequently, knowledge is inferred from *in vivo* analyses of embryos from related species as well as *in vitro* differentiation of human PSCs (hPSCs) into PGCLCs (reviewed by Kobayashi and Surani, 2018). Immediately prior to gastrulation, the pluripotent epiblast (which corresponds to PSCs) undergoes anterior-posterior patterning, generating anterior and posterior epiblast regions (Kobayashi et al., 2017; Rivera-Pérez and Magnuson, 2005). Subsequently, in mouse and pig embryos, PGCs originate in the vicinity of the posterior epiblast; the posterior epiblast also gives rise to the primitive streak (the precursor to endoderm and mesoderm, Fig. 1a) (Ginsburg et al., 1990; Kobayashi et al., 2017). This led to the hypothesis that the posterior epiblast and/or primitive

streak/mesoderm may be the precursor to human PGCs (Model 1, Fig. 1a) (Kobayashi et al., 2017; Sasaki et al., 2015). This hypothesis was initially surprising, as a hallmark of PGC development *in vivo* is "repression of somatic genes" (including primitive streak/mesodermal markers (Surani et al., 2007)). However, shared transcription factors are required for both primitive streak and PGC specification (Aramaki et al., 2013; Chen et al., 2017; Kojima et al., 2017), hinting at their intertwined origins. Alternatively, PGCs and primitive streak may be two completely independent lineages that arise from nearby precursors (Model 2, Fig. 1a). Finally, studies of cynomolgus monkey embryos have instead proposed that PGCs arise from the dorsal amnion (Sasaki et al., 2016).

Guided by these insights, prevailing strategies for hPSC differentiation into PGCLCs are divided into two phases: first, exposure to primitive streak/mesoderm-inducing signals (TGFβ and WNT) for 12–48 hours in monolayer cultures, followed by 3D aggregation and treatment with PGC-specifying signals (BMP, SCF, EGF and LIF) for multiple days (Kobayashi et al., 2017; Sasaki et al., 2015; Yokobayashi et al., 2017). This highlighted key inductive signals for PGCLC specification, although the precise temporal dynamics with which key signals (e.g., WNT) are needed remains unclear. The transcriptional network that incipiently specifies human PGCs from pluripotent cells also requires further definition. Pluripotent cells and PGCs share the expression of pluripotency transcription factors, and it is currently proposed that the pluripotency network is first abolished in differentiating cells and then reactivated in PGCs, at least in mice (reviewed by Magnusdottir and Surani, 2013; Surani et al., 2007). An alternate model is that pluripotency genes are continuously expressed from the pluripotent epiblast to the incipient PGCs, without an intermediate step where pluripotency genes are silenced. Preliminary evidence for this latter model was provided in cynomolgus monkey embryos (Sasaki et al., 2016) but it is unclear whether the same holds true for human PGC specification. Taken together, multiple outstanding questions continue to surround where, when, and how human PGCs are specified.

Here we develop a simplified 2D platform to generate human PGCLCs within 3.5 days of hPSC differentiation, and we demonstrate the applications of this simplified differentiation approach to provide new insights into PGCLC specification. Prevailing approaches to generate human PGCLCs entail 3D aggregates treated with high concentrations of growth factors (Irie et al., 2015; Kobayashi et al., 2017; Sasaki et al., 2015). While 3D aggregates may be beneficial as they concentrate intercellular interactions important for lineage specification, it is challenging to direct hPSC differentiation within 3D aggregates due to multiple difficult-to-control variables (e.g., limited diffusion of extracellular signals) (reviewed by Fowler et al., 2019). First, we adapt human PGCLC differentiation to monolayer culture and demonstrate that temporally dynamic activation, followed by repression, of WNT signaling is critical. This method can be robustly extrapolated to a wide range of hESC and hiPSC lines to achieve PGCLC specification in monolayer culture. Second, we also describe cell-surface markers for human PGCLCs (namely, CXCR4$^+$PDGFRA$^-$GARP$^-$) that enable their ready purification via fluorescence-activated cell sorting (FACS). Third, we provide a detailed characterization of cells *en route* to PGCLCs specification by single cell RNA-sequencing and find that pluripotency transcription factors OCT4 and NANOG are continuously expressed during this process, thus bridging pluripotency with germline states. Finally, single-cell RNA-

sequencing shows that monolayer culture-induced PGCLCs share strong transcriptional similarities with PGCs obtained from the human fetus.

Results

Temporally dynamic WNT activation, followed by inhibition, increases efficiency of human PGCLC specification

We hypothesized that temporal control over WNT signaling might be crucial for PGCLC specification, as the precise duration of WNT activity is of paramount importance to specify multiple cell-types from hPSCs (Loh et al., 2016). Prevailing methods for hPSC differentiation into PGCLCs generally entail two steps. First, exposure to posteriorizing signals (including TGFβ, WNT and non-specific ROCK inhibitor Y27632) that induce primitive streak/mesoderm for 12–60 hours (Kobayashi et al., 2017; Sasaki et al., 2015; Yokobayashi et al., 2017). Second, cells are then aggregated in 3D and treated with high concentrations of BMP, EGF, SCF, LIF and Y27632 for multiple days to generate PGCLCs (Kobayashi et al., 2017; Sasaki et al., 2015). Using these published protocols as a framework (**Fig. S1a**) (Kobayashi et al., 2017; Sasaki et al., 2015) and *NANOS3-mCherry* hPSCs to quantify the percentage of NANOS3[+] PGCLCs (Irie et al., 2015), we sought to examine the temporal dynamics of WNT signaling and to generate human PGCLCs in monolayer cultures.

We found that temporally dynamic activation, followed by inhibition, of WNT signaling enhanced human PGCLC specification. In the first phase of differentiation, 12 hours exposure to posteriorizing signals (including WNT agonist CHIR99021) was optimal, in order to subsequently generate NANOS3-mCherry[+] PGCLCs at the second stage of differentiation (Fig. 1b) in monolayer cultures. This reaffirmed findings that 12 hour treatment with posteriorizing signals is ideal for subsequent PGCLC differentiation (Kobayashi et al., 2017). In our hands, prolonged exposure to posteriorizing signals for 24 hours—which we and others have shown generate primitive streak (PS) cells capable of subsequent endoderm and mesoderm differentiation (Kobayashi et al., 2017; Loh et al., 2014; Loh et al., 2016)—abrogated the subsequent generation of PGCLCs in the second phase (Fig. 1b).

Subsequently, in the second phase of differentiation, we found that explicit inhibition of WNT signaling (using XAV939 (Huang et al., 2009)) led to a ~ 3-fold improvement in PGCLCs specification (Fig. 1c, compare "base condition" vs. "XAV939"). Conversely, continued WNT activation with CHIR99021 in the second phase of differentiation completely repressed PGCLC specification (Fig. 1c). The explicit requirement for WNT inhibitor (beyond simply withholding exogenous WNT) implies that differentiating hPSCs endogenously produce WNT signals (Blauwkamp et al., 2012; Loh et al., 2014; Loh et al., 2016) that inhibit PGCLCs formation. This emphasizes the need to control endogenous signals to guide efficient differentiation and is consistent with how PRDM14 inhibits endogenous WNT signaling during PGCLC specification (Sybirna et al., 2020). We conclude that temporally dynamic WNT activation followed by inhibition enhances human PGCLCs specification, thus adding a new dimension to our knowledge of

PGCLCs development; this parallels how WNT is initially required and then dispensable for pig PGCs specification in embryonic explant cultures (Kobayashi et al., 2017).

Given the importance of this initial 12-hour WNT pulse in the first phase of differentiation, we sought to molecularly detail the differentiated cells at day 0.5 (D0.5), which constitute a key intermediate *en route* to PGCLC differentiation. Single-cell RNA-sequencing (scRNAseq, using the 10X Genomics droplet-based platform (Zheng et al., 2017)) revealed that these hPSC-derived D0.5 cells are fairly homogenous, and continued to highly express pluripotency transcription factors *OCT4* and *NANOG*, although *SOX2* decreased (Fig. 1d, **Fig. S1b,c**). D0.5 cells also began to concomitantly express posterior epiblast/future primitive streak markers such as *MIXL1*, *BRACHYURY*, *FGF8* and *NODAL* (Fig. 1e, **Fig. S1b,c**). However, D0.5 cells expressed posterior epiblast/primitive streak markers at lower levels compared to D1 primitive streak cells that were generated by 24 hours exposure to posteriorizing signals (Fig. 1e). Consistent with the use of TGFβ and WNT agonists to induce D0.5 cells, these cells demonstrated an active transcriptional response to both signaling pathways, including TGFβ target genes (*FOLLISTATIN*, *ID1* and *LEFTY2*) and WNT target genes (*SP5*) (**Fig. S1b,c**).

We provisionally designate these intermediate cells generated upon 12-hour exposure to posteriorizing signals as "posterior epiblast" to distinguish them from primitive streak (PS). As discussed above, D1 PS cells are capable of generating endoderm and mesoderm, but not PGCLCs (Fig. 1b). We propose that D0.5 cells correspond to "posterior epiblast" based on how, in mouse embryos, the post-implantation pluripotent epiblast is formed by embryonic day 5.5 (E5.5), but then PS markers (e.g., Brachyury) are transiently expressed in the posterior region of the epiblast (~ E6-E6.25) immediately *prior* to overt formation of the morphologically-conspicuous PS (~ E6.5) (Rivera-Pérez and Magnuson, 2005); similar results have been reported in pig embryos (Kobayashi et al., 2017). However, we note that early human post-implantation embryos remain inaccessible for analysis, and thus assignment of terms such as "posterior epiblast" or "primitive streak" in human is premised on evolutionary homology to other mammals such as pig and mouse (Fowler et al., 2019). In summary, this discloses a unique transcriptional signature for D0.5 posterior epiblast cells, wherein pluripotency factors *OCT4* and *NANOG* are co-expressed together with primitive streak markers.

Generation of human PGCLCs in monolayer conditions

After generating presumptive posterior epiblast cells, next we tested whether continuous BMP, SCF, LIF, and EGF activation (Irie et al., 2015; Kobayashi et al., 2017; Sasaki et al., 2015) was required for the entire second phase of PGCLC differentiation in monolayers. First, omitting BMP4 from the culture media for 24 hours on day 2 of the second phase of differentiation led to a ~ 2.5-fold increase in PGCLCs specification, while the absence of SCF and EGF in the same time window was superfluous (**Fig. S2a**). Second, past 3D differentiation methods used extremely high BMP4 concentrations (200−500 ng/mL) (Irie et al., 2015; Kobayashi et al., 2017), but in our monolayer conditions, we observed that significantly lower (25-fold) BMP4 concentrations were needed (**Fig. S2b**). This is consistent with the notion that BMP signaling does not effectively diffuse through large hPSC clusters (Warmflash et al., 2014; Zhang et al., 2019). 3D aggregates may therefore impair BMP signaling, thus emphasizing potential benefits of a monolayer

differentiation system. Third, LIF, which is commonly used to enhance PGCLCs survival (Irie et al., 2015; Kobayashi et al., 2017; Sasaki et al., 2015), was dispensable in our platform (**Fig. S2c**). Fourth, we observed a peak of PGCLCs formation by day 3.5 of differentiation (**Fig. S2d**).

Combining these improvements together, we developed a monolayer, serum-free protocol (Fig. 2a) to consistently and reproducibly generate 20−30% pure NANOS3-mCherry⁺ PGCLCs within 3.5 days of *in vitro* differentiation (Fig. 2b). NANOS3-mCherry⁺ PGCLCs purified by fluorescence-activated cell sorting (FACS) expressed hallmark PGCs markers, including *POU5F1* (*OCT4*), *NANOG*, *TFCP2L1*, *PRDM1* (*BLIMP1*), *NANOS3* and *TFAP2C* (*AP2γ*) (Fig. 2c). PGCLCs generated using our protocol did not express markers of endoderm (*FOXA2*, *HHEX*) or extraembryonic fate (*CDX2*), thus reaffirming their lineage specificity (**Fig. S2e**). At the protein level, PGCLCs co-expressed NANOG, PRDM1/BLIMP1, SOX17 and OCT4/POU5F1 (Fig. 2d).

We independently validated this monolayer differentiation protocol using a separate *SOX17-GFP* knock-in reporter hPSC line (Wang et al., 2011) to track the expression of human PGCLCs marker SOX17 (Irie et al., 2015). Our monolayer differentiation protocol generated SOX17-GFP⁺ PGCLCs that expressed *NANOS3* but not endodermal marker *FOXA2* (**Fig. S2f,g**), excluding that these PGCs were SOX17⁺FOXA2⁺ endoderm (Loh et al., 2014). We ultimately applied our differentiation protocol across 5 additional hESC/hiPSC lines and found that it reproducibly generated PGCLCs (detailed below; Fig. 3).

Surface-marker profile of hPSC-derived PGCLCs: CXCR4⁺ PDGFRα⁻ GARP⁻

Current protocols to generate human PGCLCs in monolayers (this study) or aggregates (Irie et al., 2015; Kobayashi et al., 2017; Sasaki et al., 2015) generate heterogeneous cell populations containing a subset of PGCLCs; therefore cell-surface markers to selectively identify and purify PGCLCs would be a boon. EpCAM, ITGA6, PDPN, CD38, KIT and alkaline phosphatase activity have been previously reported to enrich for human or non-human primate PGCLCs (Irie et al., 2015; Sakai et al., 2020; Sasaki et al., 2015). However, at the transcriptional level, many of these markers are also expressed on undifferentiated hPSCs (**Fig. S3a**), consistent with past reports of protein expression on hPSCs (Irie et al., 2015; Sakai et al., 2020; Sasaki et al., 2015). Using our optimized monolayer platform for PGCLCs differentiation, we thus sought to discover alternative cell-surface markers to purify PGCLCs.

We robotically screened the expression of 369 cell-surface markers using high-throughput FACS (Loh et al., 2016) on *SOX17-GFP* hPSCs differentiated into D3.5 SOX17-GFP⁺ PGCLCs vs. SOX17-GFP⁻ non-PGCLCs; undifferentiated hPSCs were also included as a negative control (Fig. 3a). This confirmed that EpCAM, ITGA6, PDPN and alkaline phosphatase (Irie et al., 2015; Sakai et al., 2020; Sasaki et al., 2015) were not specific markers since they were both expressed on hPSCs as well as PGCLCs (Fig. 3a).

In our analysis, the most specific positive marker for SOX17-GFP$^+$ PGCLCs was the chemokine receptor CXCR4/CD184 (Fig. 3a), which similarly marked NANOS3-mCherry$^+$ PGCLCs (Fig. 3b). Intriguingly, in model organisms Cxcr4 is known to be expressed by PGCs (Weissman and Shizuru, 2008), where it enables their migration towards the gonads in response to Cxcl12 (Molyneaux et al., 2003); this may also be conserved in human (Mitsunaga et al., 2017). However, CXCR4 is also expressed on mesodermal derivatives (McGrath et al., 1999), and therefore negative expression of mesodermal markers is necessary to exclude mesoderm. We found that the mesodermal markers PDGFRα/CD140A (Kattman et al., 2011) and GARP/LRRC32 (Loh et al., 2016) were expressed on the D3.5 non-PGCLCs (Fig. 3a,b), thus providing a means to eliminate mesoderm.

Taken together, by relying on a combination of positive (CXCR4) and negative (PDGFRα, GARP) markers, we defined a CXCR4$^+$ PDGFRα$^-$ GARP$^-$ surface marker profile for hPSC-derived PGCLCs. Logical combinations of positive and negative markers have likewise proven decisive in the purification of specific cell-types in blood and other tissues (Weissman and Shizuru, 2008). In differentiated D3.5 cultures, the CXCR4$^+$ PDGFRα$^-$ GARP$^-$ fraction contained all PGCLCs; other combinations of these surface markers did not enrich for PGCLCs (Fig. 3c).

PGCLCs can be consistently generated across diverse hESC and hiPSC lines

Across a panel of 5 distinct, wild-type male and female hESC and hiPSC lines, the present monolayer differentiation method generated an average of 36.7±22.6% (with peaks as high as 73.2%) pure CXCR4$^+$ PDGFRα$^-$ GARP$^-$ PGCLCs (Fig. 4a). Our CXCR4$^+$ PDGFRα$^-$ GARP$^-$ cell-surface marker signature allowed us to purify differentiated PGCLCs across all hESC and hiPSC lines tested, using our improved differentiation strategy and without recourse to transgenic reporters (Fig. 4b, **Fig. S3b**). Across all lines, hPSC-derived CXCR4$^+$ PDGFRα$^-$ GARP$^-$ PGCLCs upregulated of hallmark PGC markers without substantial expression of endodermal or mesodermal markers (Fig. 4b, **Fig. S3b**). Immunostaining of FACS-sorted hPSC-derived D3.5 CXCR4$^+$ PDGFRα$^-$ GARP$^-$ cells confirmed that they ubiquitously co-expressed the PGC marker proteins SOX17, PRDM1/BLIMP1 and TFAP2C/TFAP2γ (Fig. 4c). This exemplifies the fidelity of PGCLC specification across distinct genetic backgrounds and also demonstrates the utility of the CXCR4$^+$ PDGFRα$^-$ GARP$^-$ surface marker profile.

Tracking the trajectory and uniformity of PGCLC specification in vitro **using single-cell RNA-sequencing**

Next we illuminated the stepwise changes in gene expression as hPSCs incipiently differentiated into posterior epiblast (D0.5) and then into a PGCLC-containing population (D3.5) by performing single-cell RNA sequencing (scRNAseq) of all of these populations (Fig. 5a). Given that we showed above that the D3.5 population is heterogeneous (Figs. 1−4), scRNAseq was important to detail the cellular diversity of this population and to obtain a refined and specific transcriptional signature only for the PGCLCs. As a negative control, we also performed scRNAseq of D2 definitive endoderm (Loh et al., 2014)—a lineage derived from the PS (and thus, on a related but distinct lineage path from PGCs)—to clarify the

relationship between human PGCs and endoderm, given that human PGCLCs reportedly express "endodermal" marker *SOX17* (Irie et al., 2015; Sasaki et al., 2015). Taken together, we analyzed 24,473 cells by scRNAseq, with a median of >4000 genes detected per cell in each cell population (**Fig. S4a**).

scRNAseq showed that the D3.5 bulk differentiated population was transcriptionally heterogeneous, comprising two distinct subsets (Fig. 5b, **Fig. S4b**). One subset comprised PGCLCs expressing *NANOS3*, *TFAP2C* and *KLF4* (Fig. 5b, **Fig. S4c**). Intriguingly, the non-PGCLCs expressed lateral mesoderm marker *HAND1* and the cardiac mesoderm markers *TMEM88*, *MYOSIN LIGHT CHAIN 4* (*MYL4*) and *ALPHA CARDIAC ACTIN* (*ACTC1*) (Loh et al., 2016; Novikov and Evans, 2013) (Fig. 4b, **Fig. S4c, d**). This suggests that the "mis-differentiated", non-PGCLCs at D3.5 are mesoderm-like cells, as evinced by HAND1 protein expression in the D3.5 non-PGCs (Fig. 4c). Indeed, the principal signals we used to differentiate posterior epiblast into PGCLCs (BMP activation and WNT inhibition) are the same ones that differentiate primitive streak into cardiac mesoderm (Loh et al., 2016), suggesting that some cells on the wrong differentiation trajectory respond to these same signals to adopt mesoderm-like identity.

To overcome such population heterogeneity, we asked whether our previously-described cell-surface markers (CXCR4$^+$ PDGFRα$^-$ GARP$^-$) enabled the purification of homogeneous PGCLCs. While past combinations of cell-surface markers could isolate PGCLCs that were enriched for NANOS3, PRDM1 and TFAP2C expression (Irie et al., 2015; Sasaki et al., 2015), we surmised that single-cell RNA-seq of FACS-sorted PGCLCs would rigorously assess whether they were truly homogeneous at the transcriptome-wide level. scRNAseq of FACS-sorted CXCR4$^+$ PDGFRα$^-$ GARP$^-$ D3.5 PGCLCs (as described in Fig. 3) revealed that they were essentially homogeneous (97.2% pure PGCLCs) (Fig. 5d). This result thus reaffirms the power of our cell-surface marker profile to precisely isolate PGCLCs from a heterogeneous cell population, thus opening the door to downstream functional and molecular analyses of purified PGCLCs.

Integrated scRNAseq analysis of all populations revealed the stepwise changes in gene expression as pluripotent cells segue into D0.5 posterior epiblast and, finally, D3.5 PGCLCs (Fig. 5e, **Fig. S4d**). Posterior epiblast markers *BRACHYURY*, *MIXL1* and *NODAL* were transiently expressed at D0.5 (consistent with how posterior epiblast/primitive streak transcription factors are required for mammalian PGC specification (Aramaki et al., 2013; Chen et al., 2017; Kojima et al., 2017)), but were subsequently downregulated in D3.5 PGCLCs (Fig. 5e). This is consistent with the observed "repression of somatic genes" in fully-formed PGCLCs (Surani et al., 2007), although we note that these genes are nonetheless briefly expressed in their precursors (the posterior epiblast). Our side-by-side comparison of PGCLCs and endoderm confirmed that they shared common markers *SOX17* and *PRDM1* (Irie et al., 2015; Sasaki et al., 2015); however, D3.5 PGCLCs expressed multiple unique markers that were not found in endoderm, including *NANOG*, *NANOS3*, *TFAP2C*, *KLF4*, and *TCL1B* (Fig. 5e, **Fig. S5a**), thus disclosing a single-cell transcriptional signature for hPSC-derived PGCLCs.

We then investigated expression of pluripotency markers during germline differentiation; a quintessential feature of early germ cells (unlike most somatic cell-types) is that they express pluripotency transcription factors (reviewed by Magnusdottir and Surani, 2013; Surani et al., 2007). The prevailing model is that

upon early differentiation, pluripotent cells initially downregulate pluripotency factors, but subsequently only cells allocated to the germline "re-express" pluripotency factors (reviewed by Magnusdottir and Surani, 2013; Surani et al., 2007) (**Fig. 5fi**). However, this models contrast with some recent observations made in cynomolgus macaque embryos, where NANOG was observed to be continuously expressed in PGCs precursors and PGCs; nonetheless, similar observations have been precluded in human embryos due to the lack of accessibility of proper developmental stages. To assess which of the two models may pertain to human PGCs, we computationally ordered differentiating cells in our scRNAseq dataset along an inferred "pseudotime" (Qiu et al., 2017), and observed that *POU5F1* and *NANOG* were continuously expressed during the transition from pluripotency to posterior epiblast to hPGCLCs (**Fig. 5fii**). This thus implies continuous expression of pluripotency factors in the transition from pluripotency to germline fate. We sought to experimentally validate this prediction by tracking NANOG expression at the single-cell level. To this end, we engineered *NANOG-2A-YFP* reporter hESCs wherein Cas9/AAV6 genome editing (Martin et al., 2019) was used to insert a *2A-YFP* reporter immediately downstream of the *NANOG* gene without disrupting its coding sequence (Martin et al., 2020).

NANOG was continuously expressed during the hPSC-to-germline transition, without evidence for NANOG downregulation followed by re-expression. Undifferentiated hPSCs, D0.5 posterior epiblast cells and D1.5 cells were largely NANOG$^+$ CXCR4$^-$, and by D2.5-D3.5, a subpopulation continued to express NANOG but gained CXCR4, thus transitioning to NANOG$^+$ CXCR4$^+$ PGCLCs (**Fig. 5fiii, Fig. S5b**). By contrast, by D2.5-D3.5, other cells lost NANOG, thus differentiating into NANOG$^-$ CXCR4$^-$ non-PGCLCs (**Fig. 5fiii, Fig. S5b**). We independently confirmed these results, by using intracellular flow cytometry to directly stain for NANOG protein itself (**Fig. S5b,c**).

Therefore, after the specification of NANOG$^+$ posterior epiblast, differentiating cells that "inherit" pluripotency factor expression from the posterior epiblast may progress forth to the germline through, at least in part, the inhibition of WNT signaling. Continued NANOG expression may thus serve as a bridge to link the pluripotent and PGC states. This mirrors staining analyses of cynomolgus macaque embryos (Sasaki et al., 2016), and remains to be substantiated in other species.

Single cell RNA-seq analysis confirms that in vitro-derived PGCLCs transcriptionally resemble in vivo-derived human fetal PGCs

Finally, we used single-cell RNA-seq to determine whether hPSC-derived PGCLCs resemble *bona fide* PGCs within the human fetus. Past work affirmed similarities between human PGCLCs and PGCs using bulk-population RNA-seq (Irie et al., 2015) but did not acquire single-cell resolution. Another study used single cell-RNA-seq to compare human PGCLCs and cynomolgus macaque PGCs (Chen et al., 2019), but did not compare them against human fetal PGCs.

A published scRNAseq analysis of >2000 fetal germ cells (FGCs, including a subset of PGCs) from week 5–26 human fetuses (Li et al., 2017) laid a foundation for assessing the identity of *in vitro*-derived PGCLCs since it offers a comprehensive roadmap for germ cell development *in vivo* under physiological

conditions. In that study, human FGCs were classified into four sequential subsets characterized by mitosis, retinoid signalling, meiotic prophase, and oogenesis (termed FGC1 to FGC4, respectively) (Li et al., 2017). We compared human FGCs, along with fetal gonad somatic cells (Li et al., 2017), against our hPSC-derived D3.5 PGCLCs (Fig. 6a, **Fig. S6a**).

Hierarchical clustering revealed that hPSC-derived PGCLCs were the most similar to FGC1, which represents early-stage PGCs (Fig. 6b). hPSC-derived PGCLCs and FGC1 both expressed pluripotency genes including *POU5F1* and *NANOG*, as well as PGC-specific markers such as *NANOS3*, *SOX17*, *PRDM1* and *TFAP2C* (Fig. 6c). By contrast, these pluripotency and PGCs markers were turned off in later-stage FGC2, FGC3 or FGC4 populations, consistent with exit from a PGC state *in vivo* (Fig. 6c). hPSC-derived PGCLCs appeared to represent an early PGC population, as they did not express markers of differentiating germline cells (e.g., those involved in retinoid signaling, oogenesis or meiosis), which instead were expressed in FGC2, FGC3 or FGC4 (**Fig. S6b**). While PGCLCs most closely resembled FGC1, it is not possible to access human pre-migratory PGCs at earlier developmental stages (Li et al., 2017), and it is thus possible that PGCLCs may correspond to even earlier-stage PGCs. Taken together, this shows a striking transcriptome-wide similarity between hPSC-derived PGCLCs *in vitro* and human fetal PGCs *in vivo*. hPSC-derived PGCLCs apparently represent an early PGC population prior to the initiation of germline differentiation and meiosis.

Discussion

Expanding upon past work that successfully generated PGCLCs from hPSCs in 3D aggregates (Irie et al., 2015; Kobayashi et al., 2017; Sasaki et al., 2015), here we report a simplified monolayer system to produce human PGCLCs and we exploit this system to provide new insights into PGCLC specification using single-cell RNA-sEq. Stem cells negotiate a series of branching lineage decisions during differentiation (reviewed by Fowler et al., 2019; Waddington, 1940). We hypothesized that at a critical lineage bifurcation, differentiating hPSCs may inadvertently stray down a non-PGC lineage. At this bifurcation, what signals promote PGCLC specification at the expense of non-PGCs?

A principal finding of this work is that temporally dynamic activation, followed by inhibition, of WNT signaling enhanced human PGCLC specification. In the first phase of differentiation, hPSCs are briefly exposed to primitive streak-inducing signals (WNT and TGFβ) for 12 hours to generate candidate "posterior epiblast" cells (Kobayashi et al., 2017). By single-cell RNA-seq, this intermediate cell state expresses *OCT4* and *NANOG* together with posterior epiblast/future primitive streak markers, and thus appears analogous to mouse and pig posterior epiblast cells (Kobayashi et al., 2017; Rivera-Pérez and Magnuson, 2005). In the second phase of differentiation, these posterior epiblast cells apparently face a branching lineage choice to differentiate into PGCLCs or primitive streak. At this lineage choice, continued WNT activation specified primitive streak (Kobayashi et al., 2017; Loh et al., 2014; Loh et al., 2016), and therefore WNT *inhibition* was critical to suppress primitive streak formation and to differentiate posterior epiblast into PGCLCs. In this model, the PGC and somatic (primitive streak) lineages are related—yet distinct—cell-types that both arise from posterior epiblast intermediates and are segregated by mutually-

exclusive signals (e.g., WNT). This temporally dynamic role for WNT signaling agrees with analyses of pig embryos (Kobayashi et al., 2017), but further *in vivo* analyses are warranted to confirm various aspects of this model.

At the transition from pluripotent to germline states, we find that pluripotency factor NANOG is continuously expressed in the transition from hPSCs to posterior epiblast to PGCLCs in response to WNT modulation. This mirrors preliminary evidence drawn from cynomolgus macaque embryos (Sasaki et al., 2016). It was initially hypothesized that differentiating cells lose pluripotency factor expression, but cells allocated to the germline "re-express" such factors, thus "regaining" features of pluripotency (reviewed by Magnusdottir and Surani, 2013; Surani et al., 2007). By contrast, we propose that the continued expression of pluripotency factors such as NANOG may serve as a direct molecular bridge between the pluripotent and germline states; cells that lose such expression may instead differentiate into somatic cells. Our findings are altogether consistent with Nanog's critical role in mouse PGC specification *in vivo* and *in vitro* (Chambers et al., 2007; Murakami et al., 2016).

However, in the monolayer system, typically only one-third of cells (on average 36.7%, across all hPSC lines) differentiate into PGCLCs in 3.5 days. To overcome this limitation, we discovered an alternative cell-surface marker signature for PGCLCs (CXCR4$^+$ PDGFRα$^-$ GARP$^-$) that enables their purification from multiple hPSC lines. Past markers of PGCLCs (e.g., alkaline phosphatase activity, EpCAM, ITGA6 and PDPN) were also expressed on undifferentiated hPSCs (Irie et al., 2015; Sakai et al., 2020; Sasaki et al., 2015); we find that the CXCR4$^+$ PDGFRα$^-$ GARP$^-$ signature separates PGCLCs from undifferentiated hPSCs as well as the non-PGCLCs inadvertently generated during differentiation. Our single-cell RNA-seq analysis of FACS-sorted CXCR4$^+$ PDGFRα$^-$ GARP$^-$ PGCLCs also reaffirmed that they are a nearly homogeneous population, and thus we foresee this surface-marker sorting strategy will have broad applications in isolating PGCLCs for molecular and functional experiments. Single-cell RNA-seq analysis also revealed strong transcriptome-wide similarities between hPSC-derived CXCR4$^+$ PDGFRα$^-$ GARP$^-$ PGCLCs and early human fetal PGCs (Li et al., 2017), adding to past work that demonstrated similarities by bulk-population RNA-seq (Irie et al., 2015). This thus serves to molecularly authenticate the monolayer generated PGCLCs; however, further functional *in vivo* tests await, given the ethical difficulties of transplanting human PGCLCs into animal models.

Our single-cell RNA-sequencing survey revealed that the remaining "mis-differentiated" non-PGCLCs are mesoderm-like cells. In both monolayer conditions (this study) as well as within 3D aggregates (Irie et al., 2015; Kobayashi et al., 2017; Sasaki et al., 2015), our inability to generate pure PGCLCs *in vitro* indicates that we have an incomplete understanding of the inductive and repressive extracellular signals leading to human PGCLCs specification. Recent reports suggest that repression of *Otx2* (in mouse) (Zhang et al., 2018), or overexpression of *SOX17* and *BLIMP1* (in human) (Irie et al., 2015; Kobayashi et al., 2017), suffices to generate PGCLCs *in vitro* even in the absence of any exogenous signals. Delineating the upstream extracellular signals that repress *OTX2*, or that induce *SOX17* and *BLIMP1*, is therefore paramount to further enhance the efficiency of human PGC formation *in vitro*. Collectively, our data demonstrate that it is possible to generate human PGCLCs in monolayer cultures. This platform will

simplify efforts to dissect the molecular mechanisms that regulate human PGCs induction and maturation and may accelerate the identification of culture conditions that will be conducive to the formation of fully-functional, meiosis-competent human germ cells.

Materials And Methods

Human pluripotent stem cell (hPSC) culture

H1 hESCs (Thomson et al., 1998), H9 hESCs (Thomson et al., 1998), *NANOS3-mCherry* WIS1 hESCs (Irie et al., 2015), *SOX17-GFP* H9 hESCs (Wang et al., 2011), *NANOG-2A-YFP* H9 hESCs (Martin et al., 2020), BJC1 hiPSCs (Durruthy-Durruthy et al., 2014), BJC3 hiPSCs (Durruthy-Durruthy et al., 2014), and BIRc3 hiPSCs were routinely propagated feeder-free in mTeSR1 medium + 1% penicillin/streptomycin (StemCell Technologies) on cell culture plastics coated with Matrigel (Corning). Undifferentiated hPSCs were maintained at high quality with particular care to avoid any spontaneous differentiation, which would confound downstream differentiation.

In the *NANOS3-mCherry* hESC line, a *2A-mCherry* fluorescent reporter was inserted immediately downstream of the *NANOS3* gene without disrupting its coding sequence (Irie et al., 2015). In the *SOX17-GFP* hESC line, a *GFP* fluorescent reporter was inserted immediately after the *SOX17* start codon, thus functionally invalidating one *SOX17* allele (Wang et al., 2011). In the *NANOG-2A-YFP* hESC line, Cas9 RNP/AAV6-based genome editing (Martin et al., 2019) was used to insert a *2A-iCaspase9-2A-YFP* fluorescent reporter immediately downstream of the *NANOG* gene without disrupting the *NANOG* coding sequence (Martin et al., 2020).

hPSC differentiation into PGCs

Undifferentiated hPSCs were maintained in mTeSR1 + 1% penicillin/streptomycin and enzymatically passaged (Accutase, 1:8-1:12 split) for differentiation. After overnight recovery in mTeSR1 + Thiazovivin (ROCK inhibitor, 5 µM), the following morning, hPSCs were briefly washed (DMEM/F12) and differentiated into posterior epiblast for 12 hours (100 ng/mL Activin + 3 µM CHIR99021 + 10 µM Y-27632) in aRB27 basal media, which comprised Advanced RPMI 1640 medium supplemented with 1% B27 supplement, 0.1 mM non-essential amino acids (NEAA), 100 U/ml Penicillin + 0.1 mg/ml Streptomycin, and 2 mM L-Glutamine (Kobayashi et al., 2017). Subsequently, cells were washed once more and treated with 40 ng/mL BMP4, 1 µM XAV939, and 10 µM Y-27632 for 24 hours, then 100 ng/mL SCF, 50 ng/mL EGF, 1 µM XAV939, and 10 µM Y-27632 for an additional 24 hrs, and finally 40 ng/mL BMP4, 100 ng/mL SCF, 50 ng/mL EGF, 1 µM XAV939, and 10 µM Y-27632 for 24 hours (all in aRB27 basal media). For comparison, published PGC differentiation protocols (Sasaki et al., 2015; Kobayashi et al., 2017) were performed as described previously.

The following activators and inhibitors of signaling pathways were used for differentiation:

Item name	Company	Catalog no.
Activin A	Peprotech	120-14E
BMP4	R&D Systems	314-BP-050
CHIR99021	Tocris	4423
EGF	Peprotech	AF-100-15
LIF	Peprotech	300-05
SCF	Peprotech	300-07
TC-S 7001	Tocris	4961
Thiazovivin	Tocris	3845
XAV939	Tocris	3748

Immunostaining

Cells were grown on Matrigel-coated glass coverslips (Fisher). Cells were washed with PBS and then fixed with 4% PFA (paraformaldehyde) for 15 min. Coverslips were then washed with PBS, permeabilized with 0.2% Triton X-100 and blocked with PBS-BT (3% BSA + 0.1% Triton X-100 + 0.02% sodium azide in PBS) for at least 30 min. Coverslips were incubated with primary antibodies diluted in PBS-BT overnight, and then washed with PBS-BT, subsequently incubated with secondary antibodies diluted in PBS-BT for 45 min, and then washed again. Finally, samples were mounted in ProLong Diamond anti-fade mountant with DAPI (Thermo Fisher Scientific). Images were acquired on a Leica DM4000 B (Leica Microsystems, Inc., IL, USA) equipped with a QImaging Retiga-2000R (Teledyne Photometrics, AZ, USA) digital camera using a 40X objective, and processed using FIJI (v.1.52p).

The following antibodies were used for immunostaining:

Antibody	Company	Cat. No.	Dilution
Anti-AP2g	Santa Cruz Biotech	Sc-12762	1:50
Anti-Blimp1	eBioscience	14-5963-82	1:50
Anti-Hand1	R&D Systems	AF3168	1:100
Anti-Nanog	Abcam	ab109250	1:250
Anti-Oct4	BD	611203	1:100
Anti-Sox17	R&D Systems	AF1924	1:500

High-throughput cell-surface marker screening

hESCs or differentiated PGCs were dissociated (using Accutase) and plated into individual wells of four 96-well plates, each well containing a distinct antibody against a human cell surface antigen, altogether totaling 371 unique cell-surface markers across multiple 96-well plates (LEGENDScreen PE-Conjugated Human Antibody Plates; Biolegend, 700007). For each LEGENDScreen experiment, approximately 10-70 million cells of each lineage were used. High-throughput cell-surface marker staining was largely done as per the manufacturer's recommendations, and cells were stained with a viability dye (DAPI, 1.1 µM; Biolegend) prior to analysis on a CytoFLEX Flow Cytometer (Stanford Stem Cell Institute FACS Core). Stained cells were not fixed prior to FACS analysis. Sometimes, after lysophilized antibodies were reconstituted in LEGENDScreen plates they were aliquoted into a separate plate to generate replicates of

antibody arrays. Undifferentiated H9 hESCs and *SOX17-GFP* H9 hESCs were used for LEGENDScreen analyses.

Flow cytometry and fluorescence-activated cell sorting (FACS) for cell-surface marker expression

hPSCs or their differentiated derivatives were dissociated using TrypLE Express (Gibco), were washed off the plate with FACS buffer (PBS + 0.1% BSA fraction V [Gibco] + 1 mM EDTA [Gibco] + 1% penicillin/streptomycin [Gibco]) and were pelleted by centrifugation (5 mins, 4 °C). Subsequently, cell pellets were directly resuspended in FACS buffer containing pre-diluted primary antibodies (listed below), thoroughly triturated to ensure a single cell suspension, and primary antibody staining was conducted for 30 mins on ice. Afterwards, cells were washed with an excess of FACS buffer and pelleted again, and this was conducted one more time. Finally, washed cell pellets were resuspended in FACS buffer containing 1.1 µM DAPI (Biolegend), and were strained through a 35 µm filter. Flow cytometry and sorting was conducted on a BD FACSAria II (Stanford Stem Cell Institute FACS Core).

The following antibodies were used for flow cytometry:

Antibody	Company	Catalog Number	Dilution
CXCR4 APC	BD Biosciences	560936	1:5
GARP FITC	eBioscience	11-9882-41	1:20
PDGFRα PE	BD Biosciences	556002	1:20

RNA extraction and reverse transcription

In general, RNA was extracted from undifferentiated or differentiated hPSC populations plated in 12-well format by lysing them with 350 µL RLT Plus Buffer per well. RNA was extracted with the RNeasy Plus Mini Kit (Qiagen) as per the manufacturer's instructions. Generally 50-200 ng of total RNA was reverse-transcribed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) to generate cDNA libraries for qPCR.

Quantitative PCR (qPCR)

Total cDNA was diluted 1:10-1:30 in H_2O and qPCR was performed with the SensiFAST SYBR Hi-ROX Kit (Bioline) with 10 µL qPCR reactions per well in a 384-well plate: each individual reaction contained 5 µL 2x SensiFAST SYBR qPCR Master Mix + 4.2 µL cDNA (totaling ~120 ng of cDNA) + 0.8 µL of 10 µM primer stock (5 µM forward + 5 µM reverse primers). In general, gene-specific primer pairs for qPCR were tested for (1) specificity of amplicon amplification (only one peak on a dissociation curve) and (2) linearity of amplicon amplification (linear detection of gene expression in cDNA samples serially diluted seven times over two orders of magnitude, with 90-110% efficiency of amplification deemed acceptably linear). After qPCR plates were prepared by arraying sample-specific cDNAs and gene-specific primers

(listed below), they were sealed and briefly centrifuged (5 mins). 384-well qPCR plates and their adhesive sealing sheets were obtained from Thermo (AB1384 and AB0558, respectively). qPCR plates were run on a 7900HT Fast Real-Time PCR System (Applied Biosystems) with the following cycling parameters: initial dissociation (95 °C, 2 mins) followed by 40 cycles of amplification and SYBR signal detection (95 °C dissociation, 5 seconds; 60 °C annealing, 10 seconds; followed by 72 °C extension, 30 seconds), with a final series of steps to generate a dissociation curve at the end of each qPCR run. During qPCR data analysis, the fluorescence threshold to determine Ct values was set at the linear phase of amplification.

The following primers were used for qPCR analysis:

Gene Name	Forward	Reverse
BRACHYURY	TGCTTCCCTGAGACCCAGTT	GATCACTTCTTTCCTTTGCATCAAG
CDX2	GGGCTCTCTGAGAGGCAGGT	CCTTTGCTCTGCGGTTCTG
KIT	GGATTCCCAGAGCCCACAA	ACATCCACTGGCAGTACAGAA
FOXA2	GGGAGCGGTGAAGATGGA	TCATGTTGCTCACGGAGGAGTA
FZD8	ATCGGCTACAACTACACCTACA	GTACATGCTGCACAGGAAGAA
GATA4	TCCCTCTTCCCTCCTCAAAT	TCAGCGTGTAAAGGCATCTG
HHEX	CACCCGACGCCCTTTTACAT	GAAGGCTGGATGGATCGGC
MIXL1	GGTACCCCGACATCCACTTG	TAATCTCCGGCCTAGCCAAA
NANOG	AGAACTCTCCAACATCCTGAACCTC	CTGAGGCCTTCTGCGTCACA
NANOS3	ACTTACTGGCCAGGGCTACAC	ACTTCCCGGCACCTCTGAA
OCT4 (POU5F1)	AGTGAGAGGCAACCTGGAGA	ACACTCGGACCACATCCTTC
PAX5	AAACCAAAGGTCGCCACAC	GTTGATGGAACTGACGCTAGG
PRDM1	TCTCCAATCTGAAGGTCCACCTG	GATTGCTGGTGCTGCTAAATCTCTT
SHISA2	TTCCTTTACTGAAGGGAGACGAAGG	CCATCCAAAGGAATCGTGCCATAAA
SOX17	CGCACGGAATTTGAACAGTA	GGATCAGGGACCTGTCACAC
SOX2	TGGACAGTTACGCGCACAT	CGAGTAGGACATGCTGTAGGT
TFAP2C	ATTAAGAGGATGCTGGGCTCTG	CACTGTACTGCACACTCACCTT
TFCP2L1	AGCTCAAAGTTGTCCTACTGCC	TTCTAACCCAAGCACAGATCCC
VASA (DDX4)	GTGCCCTATGTGCCGTTAC	GGCTGACGTTGGACTGAGG

Single cell RNA-sequencing

We performed single-cell RNA-sequencing of undifferentiated H9 hPSCs as well as differentiated D0.5 posterior epiblast, D3.5 PGCLCs (bulk population), D3.5 FACS-sorted CXCR4$^+$ PDGFRA$^-$ GARP$^-$ PGCLCs, and D2 definitive endoderm populations. Cells of various stages were dissociated and washed twice in wash buffer (0.04% Bovine Serum Albumin in Ca2$^+$/Mg^{2+}-free PBS) and counted on the Countess II automated cell counter (Thermo Fisher). For each cell population, 10,400 cells were loaded per lane on the 10x Genomics Chromium platform, with the goal of capturing 6,000 cells. Cells were then processed for cDNA synthesis and library preparation using 10X Genomics Chromium Version 2 chemistry (catalog number 120234) as per the manufacturer's protocol. cDNA libraries were checked for quality on the

Agilent 4200 Tape Station platform and their concentration was quantified by KAPA qPCR. Libraries were sequenced using a HiSeq 4000 (Illumina) to a depth of, at a minimum, 70,000 reads per cell.

Single cell RNA-seq computational analysis of hPSC-derived cell-types

Illumina base call files were converted to FASTQ files using the Cell Ranger v2.0 program. FASTQ files were then aligned to the hg19 human reference genome using Cell Ranger. The Seurat R package (v2.3.1) (Butler et al., 2018) was used for subsequent analyses. Cells from all the various timepoints were first combined into a single Seurat object. For quality control, we first filtered out low-quality cells that expressed fewer than 2,500 genes; we also excluded cells that expressed more than 7,500 genes (which would imply doublets) or that expressed more than 0.15% mitochondrial genes (indicative of dead cells in this dataset). Counts were normalized and scaled by a factor of 10,000. To adjust for cell cycle effects, S phase and G2M genes were regressed out before Principal Component Analysis (PCA) was performed using the highly-variable genes.

For further analyses, 1,000 cells were randomly sampled from each of the 5 data sets (D0, D0.5, D3.5 sorted, D3.5 unsorted and D2 definitive endoderm) and then combined into a new file for further analysis. The top six principal components were used for clustering using the Shared Nearest Neighbor (SNN) algorithm, which was implemented via the FindCluster function in Seurat (Butler et al., 2018). Clusters were visualized in t-SNE dimensional reduction plots with 3-dimensional embedding. Differentially expressed genes between clusters were identified using the Wilcoxon rank sum test, which was performed via the Seurat package (Butler et al., 2018). For all other independent library analyses, the following numbers of principal components were used: Day 3.5 sorted library (15 principal components); Day 3.5 unsorted library (10 principal components); Day 0 vs. Day 0.5 analysis (10 principal components).

Starting from genes that were differentially expressed between each cell-type, we specifically discovered transcription factors (TFs (Chawla et al., 2013)), cell-surface proteins (Bausch-Fluck et al., 2015), and signaling ligands and receptors (Graeber and Eisenberg, 2001) whose expression was enriched in each cell population, using published and curated lists for each set of genes.

All single-cell RNA-seq plots were generated using Seurat (Butler et al., 2018) and ggplot2 (https://ggplot2.tidyverse.org/) R packages.

Finally, we computationally inferred a trajectory for progression from D0 hPSCs to D0.5 posterior epiblast to the D3.5 PGCLC-containing population (**Fig. 4f**). For this analysis, we used the whole, unsorted D3.5 population, which contains both PGCLCs and non-PGCLCs (i.e., mesoderm-like cells) in order to capture the divergence between these two mutually-exclusive lineages. For trajectory inference, we only used genes expressed in at least 10 cells. The top 1000 differentially expressed genes between these 3 cell populations (D0, D0.5 and D3.5) were used for pseudotemporal ordering and dimensional reduction was done with DDRTree using the Monocle version 2 package (Qiu et al., 2017).

Single cell RNA-seq computational analysis of human fetal germ cells

Single-cell RNA-seq data of 2629 human fetal gonadal cells *in vivo*—including both fetal germ cells (FGCs) and gonadal somatic cells—were previously published (Li et al., 2017) and were downloaded from Gene Expression Omnibus (GSE86146). Using the Seurat v3 platform (Stuart et al., 2019), we performed quality control pre-processing on this *in vivo* dataset to *1*) filter out genes that were expressed in fewer than 10 cells, *2*) exclude low-quality cells that expressed fewer than 2000 genes, and *3*) exclude low-quality cells with mitochondrial gene content greater than 5%; after these quality control steps, we obtained 2321 high-quality cells that we used for the following analysis.

Since our goal was to compare *in vivo* FGCs against *in vitro* hPSC-derived PGCLCs (and we used the female H9 hPSC line for our single-cell RNA-seq studies), we then selected female FGCs from the *in vivo* dataset (Li et al., 2017). A total of 1180 female *in vivo* cells were available and used; 732 were FGCs with ages spanning from 4 weeks to 24 weeks of fetal life, while the remaining 448 were gonadal somatic cells. The study that originally reported this *in vivo* FGC dataset classified these FGCs into four transcriptional subclusters (FGC1, FGC2, FGC3 and FGC4) (Li et al., 2017), and in our study, we used the same subclusters for transcriptional comparisons against hPSC-derived PGCLCs.

To compare FGCs against *in vitro* hPSC-derived PGCLCs, we applied the same quality control pre-processing on the FACS-sorted CXCR4$^+$ PDGFRA$^-$ GARP$^-$ PGCLC single-cell RNA-seq dataset (with 5447 high-quality cells obtained from the original dataset of 5467 cells). We randomly selected 300 of these hPSC-derived FACS-sorted PGCLCs for downstream analysis, and showed that such random downsampling of the PGCLC population did not substantially affect data quality (**Fig. S6a**). We then integrated the *in vivo* FGC and *in vitro* hPSC-derived PGCLC single-cell RNA-seq datasets using SCTransform (Hafemeister and Satija, 2019).

We obtained the list of differentially expressed genes between the FGC1, FGC2, FGC3 and FGC4 populations *in vivo* from the original publication (Li et al., 2017), with the exception that we applied a slightly more stringent statistical threshold to identify differentially expressed genes (power > 0.5). For these genes that were differentially expressed between FGC1, FGC2, FGC3 and FGC4, we also quantified the expression levels of such genes in the hPSC-derived PGCLC population. This combined expression table of *in vivo* vs. *in vitro* cells is provided in **Table S1**, and expression values are presented as transcripts per million (TPM) and $\log_e$ transformation was applied. To determine the similarity in gene expression profiles between hPSC-derived PGCLCs *in vitro* and FGC1, FGC2, FGC3 and FGC4 *in vivo*, we performed hierarchical clustering using the Average Linkage method and using a Euclidean distance metric, which showed that FGC1 and hPSC-derived PGCLCs clustered together. Pairwise Pearson's correlation coefficient was calculated for hPSC-derived PGCLCs *in vitro* and FGC1, FGC2, FGC3 and FGC4 *in vivo* using the GGally (1.5.0) R package.

Declarations

DATA AVAILABILITY

All raw data will be available upon request. RNA-seq data have been deposited to GEO. Once accession numbers are available, they will be provided to reviewers. Accession numbers will be provided here before publication.

References

1. Aramaki, S., Hayashi, K., Kurimoto, K., Ohta, H., Yabuta, Y., Iwanari, H., Mochizuki, Y., Hamakubo, T., Kato, Y., Shirahige, K., *et al.* (2013). A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. Developmental Cell *27*, 516–529.

2. Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A.P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R.L., Yoon, C., *et al.* (2015). A Mass Spectrometric-Derived Cell Surface Protein Atlas. PLoS ONE *10*, e0121314-0121322.

3. Blauwkamp, T.A., Nigam, S., Ardehali, R., Weissman, I.L., and Nusse, R. (2012). Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. Nat Commun *3*, 1070.

4. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology *36*, 411–420.

5. Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline

development. Nature *450*, 1230–1234.

6. Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A., and Kuiper, M. (2013). TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. Bioinformatics *29*, 2519–2520.

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

8. Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W., *et al.* (2019). Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. Cell Reports *29*, 4568–4582.e4565.

9. Chiquoine, A.D. (1954). The identification, origin, and migration of the primordial germ cells in the mouse embryo. The Anatomical Record *118*, 135–146.

10. De Felici, M. (2016). The Formation and Migration of Primordial Germ Cells in Mouse and Man. In Results and problems in cell differentiation (Cham: Springer International Publishing), pp. 23–46.

11. Durruthy-Durruthy, J., Briggs, S.F., Awe, J., Ramathal, C.Y., Karumbayaram, S., Lee, P.C., Heidmann, J.D., Clark, A., Karakikes, I., Loh, K.M., *et al.* (2014). Rapid and efficient conversion of integration-free human induced pluripotent stem cells to GMP-grade culture conditions. PLoS ONE *9*, e94231.

12. Fang, F., Li, Z., Zhao, Q., Ye, Z., Gu, X., Pan, F., Li, H., Xiang, W., and Xiong, C. (2020). Induced Pluripotent Stem Cells Derived From Two Idiopathic Azoospermia Patients Display Compromised Differentiation Potential for Primordial Germ Cell Fate. Frontiers in cell and developmental biology *8*, 1379.

13. Fowler, J.L., Ang, L.T., and Loh, K.M. (2019). A critical look: Challenges in differentiating human pluripotent stem cells into desired cell types and organoids. Wiley Interdisciplinary Reviews: Developmental Biology *113*, 891–823.

14. Ginsburg, M., Snow, M.H., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. Development *110*, 521–528.

15. Graeber, T.G., and Eisenberg, D. (2001). Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. Nature Genetics *29*, 295–300.

16. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biology, 1–15.

17. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519–532.

18. Huang, S.-M.A., Mishina, Y.M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G.A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., *et al.* (2009). Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. Nature *461*, 614–620.

19. Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. Cell, 1–26.

20. Kattman, S.J., Witty, A.D., Gagliardi, M., Dubois, N.C., Niapour, M., Hotta, A., Ellis, J., and Keller, G. (2011). Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell *8*, 228–240.

21. Kobayashi, T., and Surani, M.A. (2018). On the origin of the human germline. Development *145*, dev150433.

22. Kobayashi, T., Zhang, H., Tang, W.W.C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S., Contreras, D.A., Webb, R., *et al.* (2017). Principles of early human development and germ cell program from conserved model systems. Nature, 1–21.

23. Kojima, Y., Sasaki, K., Yokobayashi, S., Sakai, Y., Nakamura, T., Yabuta, Y., Nakaki, F., Nagaoka, S., Woltjen, K., Hotta, A., *et al.* (2017). Evolutionarily Distinctive Transcriptional and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. Cell Stem Cell *21*, 517–532.e515.

24. Lawson, K.A., and Hage, W.J. (1994). Clonal analysis of the origin of primordial germ cells in the mouse. Ciba Foundation Symposium *182*, 68-84- discussion 84–91.

25. Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X., *et al.* (2017). Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. Cell Stem Cell *20*, 858–873.e854.

26. Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y.Y., Nichane, M., *et al.* (2014). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell Stem Cell *14*, 237–252.

27. Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M., *et al.* (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451–467.

28. Magnusdottir, E., and Surani, M.A. (2013). How to make a primordial germ cell. Development *141*, 245–252.

29. Martin, R.M., Fowler, J.L., Cromer, M.K., Lesch, B.J., Ponce, E., Uchida, N., Nishimura, T., Porteus, M.H., and Loh, K.M. (2020). Improving the safety of human pluripotent stem cell therapies using genome-edited orthogonal safeguards. Nature Communications, 1–14.

30. Martin, R.M., Ikeda, K., Cromer, M.K., Uchida, N., Nishimura, T., Romano, R., Tong, A.J., Lemgart, V.T., Camarena, J., Pavel-Dinu, M., *et al.* (2019). Highly Efficient and Marker-free Genome Editing of Human Pluripotent Stem Cells by CRISPR-Cas9 RNP and AAV6 Donor-Mediated Homologous Recombination. Cell Stem Cell *24*, 821–828.e825.

31. McGrath, K.E., Koniski, A.D., Maltby, K.M., McGann, J.K., and Palis, J. (1999). Embryonic expression and function of the chemokine SDF-1 and its receptor, CXCR4. Developmental Biology *213*, 442–456.

32. Mitsunaga, S., Odajima, J., Yawata, S., Shioda, K., Owa, C., Isselbacher, K.J., Hanna, J.H., and Shioda, T. (2017). Relevance of iPSC-derived human PGC-like cells at the surface of embryoid bodies to prechemotaxis migrating PGCs. Proceedings of the National Academy of Sciences of the United States of America *114*, E9913-E9922.

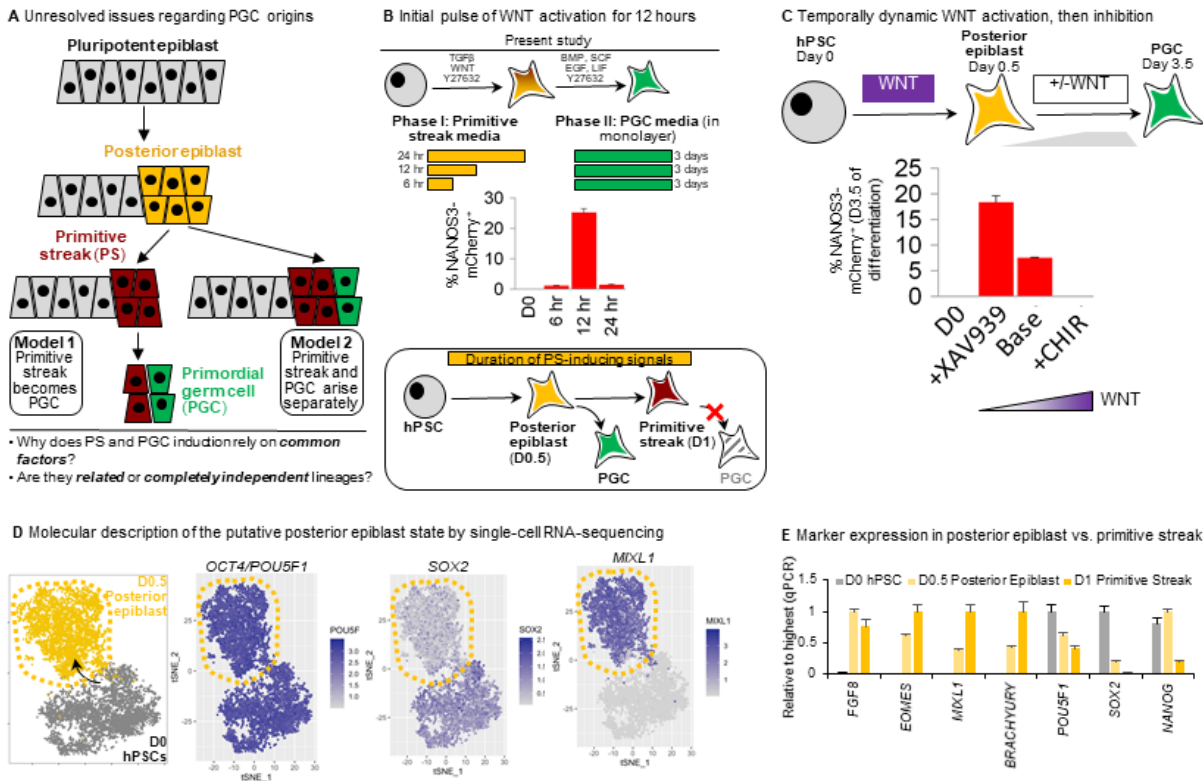33. Molyneaux, K.A., Zinszner, H., Kunwar, P.S., Schaible, K., Stebler, J., Sunshine,M.J., O'Brien, W., Raz, E., Littman, D., Wylie, C., *et al.* (2003). The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. Development*130*, 4279–4286.

34. Murakami, K., Günesdogan, U., Zylicz, J.J., Tang, W.W.C., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S., and Surani, M.A. (2016). NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. Nature *529*, 403–407.

35. Novikov, N., and Evans, T. (2013). Tmem88a mediates GATA-dependent specification of cardiomyocyte progenitors by restricting WNT signaling. Development *140*, 3787–3798.

36. O'Rahilly, R., and Müller, F. (1987). Developmental stages in human embryos (Baltimore, Maryland: Carnegie Institute of Washington).

37. Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. Cell *137*, 571–584.

38. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nature Methods *14*, 979–982.

39. Rivera-Pérez, J.A., and Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. Developmental Biology *288*, 363–371.

40. Sakai, Y., Nakamura, T., Okamoto, I., Gyobu-Motani, S., Ohta, H., Yabuta, Y., Tsukiyama, T., Iwatani, C., Tsuchiya, H., Ema, M., *et al.* (2020). Induction of the germ cell fate from pluripotent stem cells in cynomolgus monkeys. Biology of Reproduction *102*, 620–638.

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

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

43. Sosa, E., Chen, D., Rojas, E.J., Hennebold, J.D., Peters, K.A., Wu, Z., Lam, T.N., Mitchell, J.M., Sukhwani, M., Tailor, R.C., *et al.* (2018). Differentiation of primate primordial germ cell-like cells following transplantation into the adult gonadal niche. Nature Communications *9*, 5339–5313.

44. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e1821.

45. Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. Cell *128*, 747–762.

46. Sybirna, A., Tang, W.W.C., Pierson Smela, M., Dietmann, S., Gruhn, W.H., Brosh, R., and Surani, M.A. (2020). A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrons. Nature Communications *11*, 1282–1218.

47. Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–

1147.

48. Waddington, C.H. (1940). Organisers and Genes (Cambridge, UK: Cambridge University Press).

49. Wang, P., Rodriguez, R.T., Wang, J., Ghodasara, A., and Kim, S.K. (2011). Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. Cell Stem Cell *8*, 335–346.

50. Warmflash, A., Sorre, B., Etoc, F., Siggia, E.D., and Brivanlou, A.H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nature Methods *11*, 847–854.

51. Weissman, I.L., and Shizuru, J.A. (2008). The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. Blood *112*, 3543–3553.

52. Yokobayashi, S., Okita, K., Nakagawa, M., Nakamura, T., Yabuta, Y., Yamamoto, T., and Saitou, M. (2017). Clonal Variation of Human Induced Pluripotent Stem Cells for Induction into the Germ Cell Fate. Biology of Reproduction.

53. Zhang, J., Zhang, M., Acampora, D., Vojtek, M., Yuan, D., Simeone, A., and Chambers, I. (2018). OTX2 restricts entry to the mouse germline. Nature *562*, 595–599.

54. Zhang, Z., Zwick, S., Loew, E., Grimley, J.S., and Ramanathan, S. (2019). Mouse embryo geometry drives formation of robust signaling gradients through receptor localization. Nature Communications *10*, 4516–4514.

55. Zhao, Y., Ye, S., Liang, D., Wang, P., Fu, J., Ma, Q., Kong, R., Shi, L., Gong, X., Chen, W., *et al.* (2018). In Vitro Modeling of Human Germ Cell Development Using Pluripotent Stem Cells. Stem Cell Reports *10*, 509–523.

56. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., *et al.* (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications *8*, 14049.

Figures

Fig. 1

A Unresolved issues regarding PGC origins

Pluripotent epiblast

Posterior epiblast

Primitive streak (PS)

Model 1 Primitive streak becomes PGC

Model 2 Primitive streak and PGC arise separately

Primordial germ cell (PGC)

- Why does PS and PGC induction rely on *common factors*?
- Are they *related* or *completely independent* lineages?

B Initial pulse of WNT activation for 12 hours

Present study

TGFβ WNT Y27632 → BMP, SCF EGF, LIF Y27632 →

Phase I: Primitive streak media

24 hr
12 hr
6 hr

Phase II: PGC media (in monolayer)

3 days
3 days
3 days

% NANOS3-mCherry⁺: D0, 6 hr, 12 hr, 24 hr

Duration of PS-inducing signals

hPSC → Posterior epiblast (D0.5) → Primitive streak (D1) ✗

PGC   PGC

C Temporally dynamic WNT activation, then inhibition

hPSC Day 0 → WNT → Posterior epiblast Day 0.5 → +/-WNT → PGC Day 3.5

% NANOS3-mCherry⁺ (D3.5 of differentiation): D0 +XAV939, Base, +CHIR

WNT

D Molecular description of the putative posterior epiblast state by single-cell RNA-sequencing

D0.5 Posterior epiblast

D0 hPSCs

OCT4/POU5F1   POU5F
SOX2   SOX2
MIXL1   MIXL1

E Marker expression in posterior epiblast vs. primitive streak

Relative to highest (qPCR)

■ D0 hPSC ■ D0.5 Posterior Epiblast ■ D1 Primitive Streak

FGF8, EOMES, MIXL1, BRACHYURY, POU5F1, SOX2, NANOG

Figure 1

Temporally dynamic WNT activation followed by inhibition promotes human PGCLC formation a. Schematic of proposed steps of PGC development in early embryogenesis and biological questions. In Model 1, primitive streak/mesoderm-like cells give rise to PGCs (Sasaki et al., 2015). In Model 2, primitive streak and PGCs arise separately. Depicted cell positions are based on pig embryos (Kobayashi et al., 2017). b. Exposure to primitive streak-inducing signals for 12 hours is optimal for subsequent PGCLC

specification; NANOS3-mCherry hESCs were exposed to primitive streak-inducing signals (Activin + CHIR + Y27632) for either 6, 12, or 48 hours (phase I), and then transferred to PGCLC-specifying media for 3 days (phase II), and flow cytometry was then performed c. WNT inhibition promotes differentiation of posterior epiblast into PGCLCs; NANOS3-mCherry hESCs were differentiated into posterior epiblast using a WNT agonist (12 hours, phase I), and then were transferred into PGCLC-specifying media for 3 days in the presence of WNT agonist (CHIR99021) or WNT inhibitor (XAV939) (phase II); flow cytometry was then performed on D3.5 d. t-SNE plot of single-cell RNA-sequencing of posterior epiblast cells (D0.5 of differentiation) and hPSCs (D0 of differentiation) showing expression of pluripotency markers OCT4/POU5F1 and SOX2 and primitive streak marker MIXL1 e. qPCR analysis of D0 (hPSC), D0.5 (posterior epiblast) and D1 (primitive streak) of differentiation showing expression of pluripotency and primitive streak markers; qPCR data were normalized to the sample with the highest expression (which was set = 1.0); error bars = standard error

Fig. 2

A New signaling strategy: monolayer PGC induction
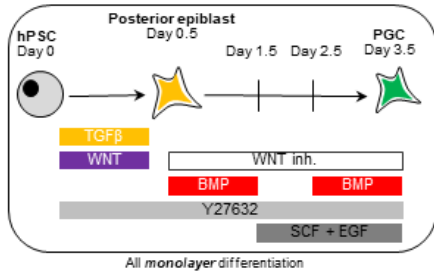
B Generation of NANOS3⁺ PGCs in monolayer conditions
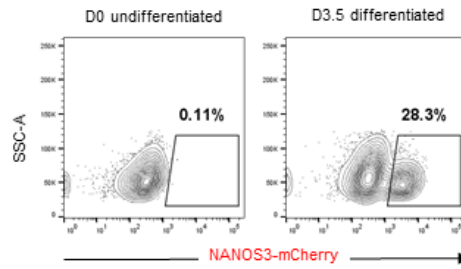
C hPSC-derived NANOS3⁺ PGCs express PGC markers
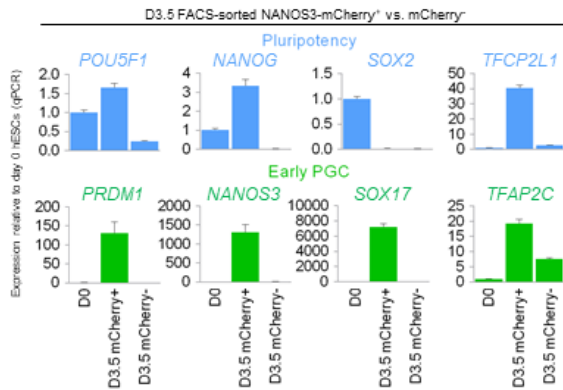
D hPSC-derived D3.5 differentiated populations express PGC marker proteins
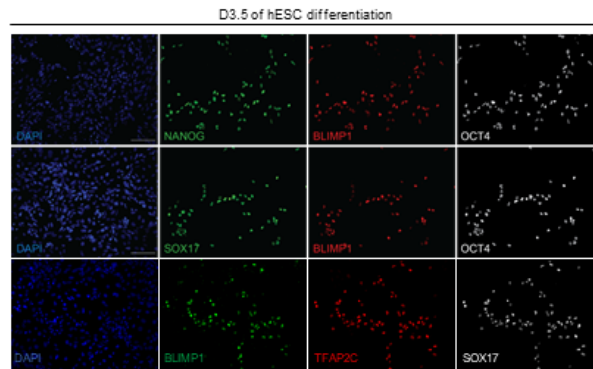
Figure 2

Human PGCLC specification in simplified monolayer cultures a. Schematic of the 2D monolayer PGCLC differentiation protocol reported in this manuscript b. Flow cytometry analysis of NANOS3-mCherry hESC shows fluorescent reporter expression before or after 3.5 days of differentiation c. qPCR analysis of NANOS3-mCherry+ PGCLCs and NANOS3-mCherry- non-PGCLCs derived after 3.5 days of differentiation, as shown in (h); as a negative control, undifferentiated hPSCs (D0) were also analyzed, and gene

expression is shown relative to undifferentiated hPSCs (which was set = 1.0) d. Immunostaining of hPSCs differentiated for D3.5 showing expression of PGC markers in a subset of cells (nuclear counterstain: DAPI). Scale bar = 100µm.

Fig. 3

A Surface marker screen of PGCs, non-PGCs and undifferentiated hPSCs

B hPSC-derived NANOS3+ PGCs are CXCR4+ GARP- PDGFRα-

C CXCR4+ GARP- PDGFRα- surface marker profile enables purification of hESC-derived PGCs without genetic reporters

qPCR of FACS-sorted D3.5 populations
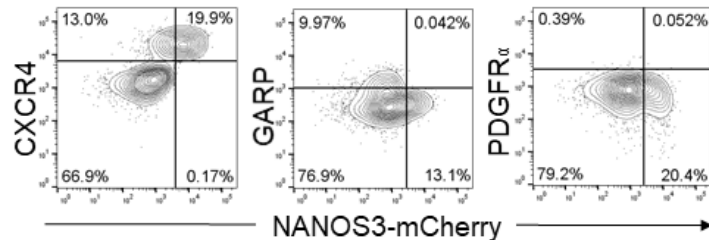
Figure 3

High-throughput screening identifies a CXCR4+ PDGFRA- GARP- cell-surface marker signature for hPSC-derived PGCLCs a. Heatmap of surface markers expressed in undifferentiated hPSC (D0), D3.5 SOX17-

GFP+ PGCLCs and D3.5 SOX17-GFP- non-PGCLCs identified from LEGENDScreen; to discriminate PGCLCs vs. non-PGCLCs, SOX17-GFP hESCs were differentiated for D3.5 and then subgated on GFP+ and GFP- before further analysis of surface marker expression; color shades represent the percentage of cells in each expression that are positive for a given marker; each row depicts expression of a single surface marker across all populations b. Flow cytometry analysis of D3.5 differentiated NANOS3-mCherry hESCs reveals CXCR4, GARP and PDGFRα expression relative to NANOS3-mCherry fluorescent reporter expression c. Flow cytometry gating strategy to identify CXCR4+/GARP-/PDGFRα- PGCLCs derived from H9 hESCs (that did not carry any fluorescent reporters) that were differentiated for D3.5; various cell populations from the D3.5 population were FACS sorted and subject to qPCR analysis, revealing that pluripotency and PGC markers are restricted to the CXCR4+/GARP-/ PDGFRα- subset and therefore reaffirming its PGCLC identity

Fig. 4

A Monolayer differentiation of PGCLCs across hPSC lines
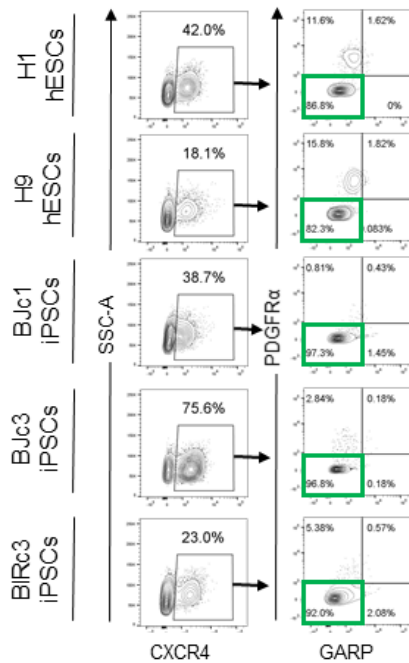
C Validation of FACS-purified CXCR4+ GARP- PDGFRA- PGCs
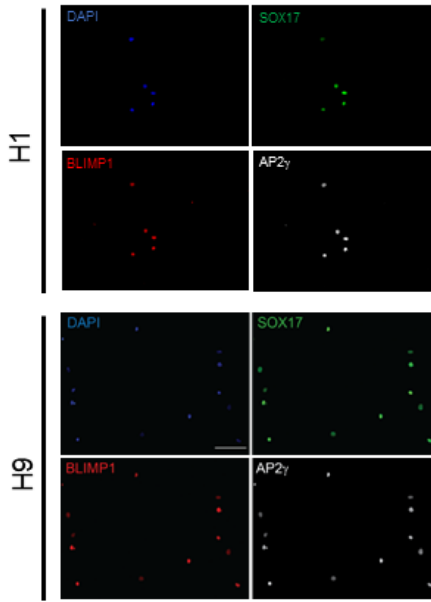
B Reproducible gene expression of PGCLCs generated from multiple hPSC lines
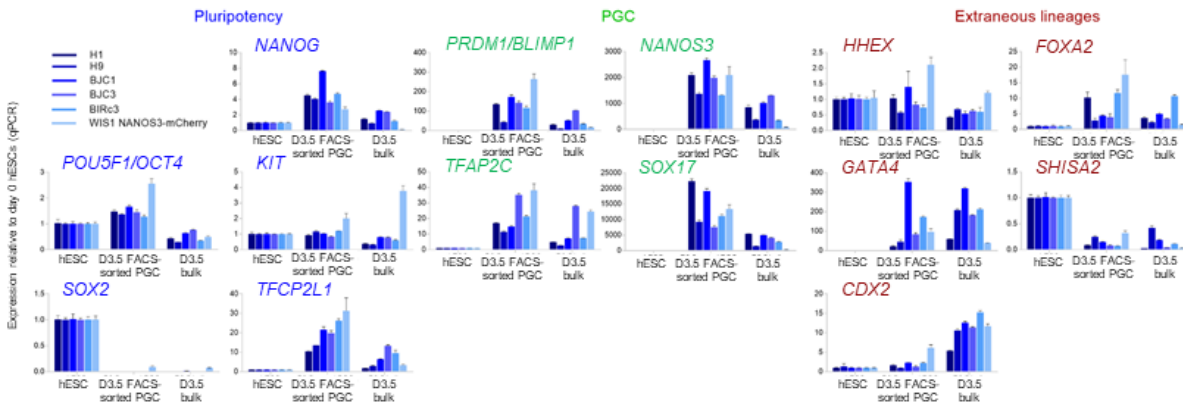
Figure 4

Reproducibility of PGCLC induction protocol across multiple hPSC lines and comparison against published differentiation protocols a. Multiple hESC/hiPSC lines (H1, H9, BJC1, BJC3, BIRc3) were differentiated into PGCLCs using our monolayer protocol and then flow cytometry analysis to quantify the purity of CXCR4+/GARP-/PDGFRα- PGCLCs was performed. b. To confirm that PGCLC differentiation is robust across multiple genetic backgrounds, multiple hESC lines (H1, H9, WIS1 NANOS3-mCherry) and

hiPSC lines (BJC1, BJC3, BIRc3), were differentiated for D3.5, and then qPCR was performed on either the bulk D3.5 population or the FACS-purified D3.5 PGCLC population (which FACS sorted for CXCR4+/GARP-/ PDGFRα- cells in the H1, H9, BJC1, BJC3, BIRc3 lines, or for NANOS3-mCherry expression in the NANOS3-mCherry line). qPCR data were normalized to expression in undifferentiated cells (which was set = 1.0). c. Immunostaining validation of FACS-purified CXCR4+/GARP-/ PDGFRα- D3.5 hPSC-derived PGCLCs (from H1 male and H9 female hESCs) confirms that they ubiquitously express PGC marker proteins (nuclear counterstain: DAPI). Scale bar = 100 µm.

Fig. 5

A Single-cell transcriptional analysis of germ cell differentiation
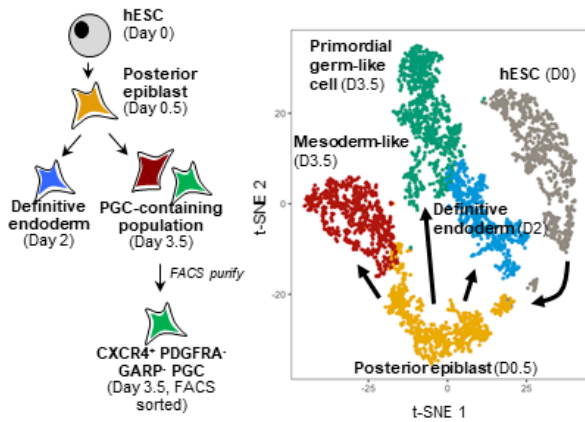
B D3.5 bulk population is heterogeneous, comprises PGCs and mesoderm-like cells
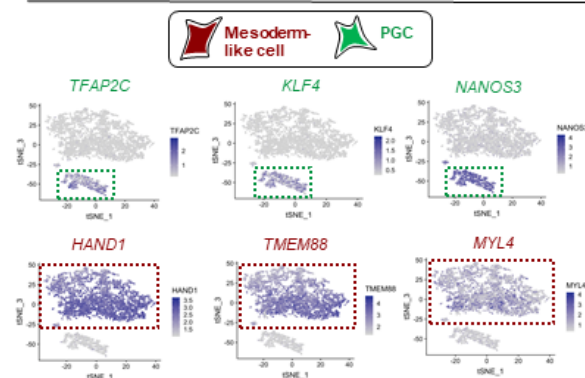
C D3.5 non-PGCs are mesoderm-like cells
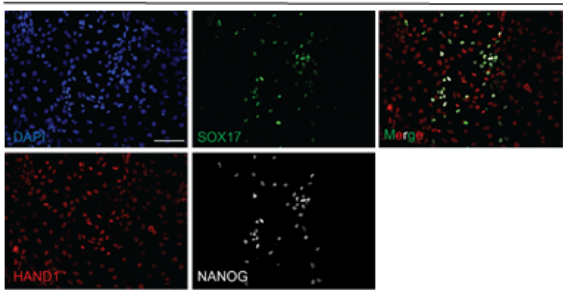
D CXCR4+ GARP- PDGFRA- FACS sorting isolates nearly-pure PGCs
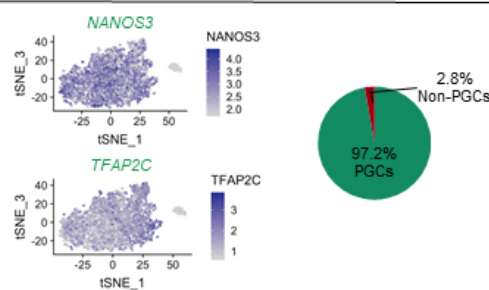
E Marker expression and cellular diversity across germ cell differentiation and in endoderm cells by single-cell RNA-sequencing
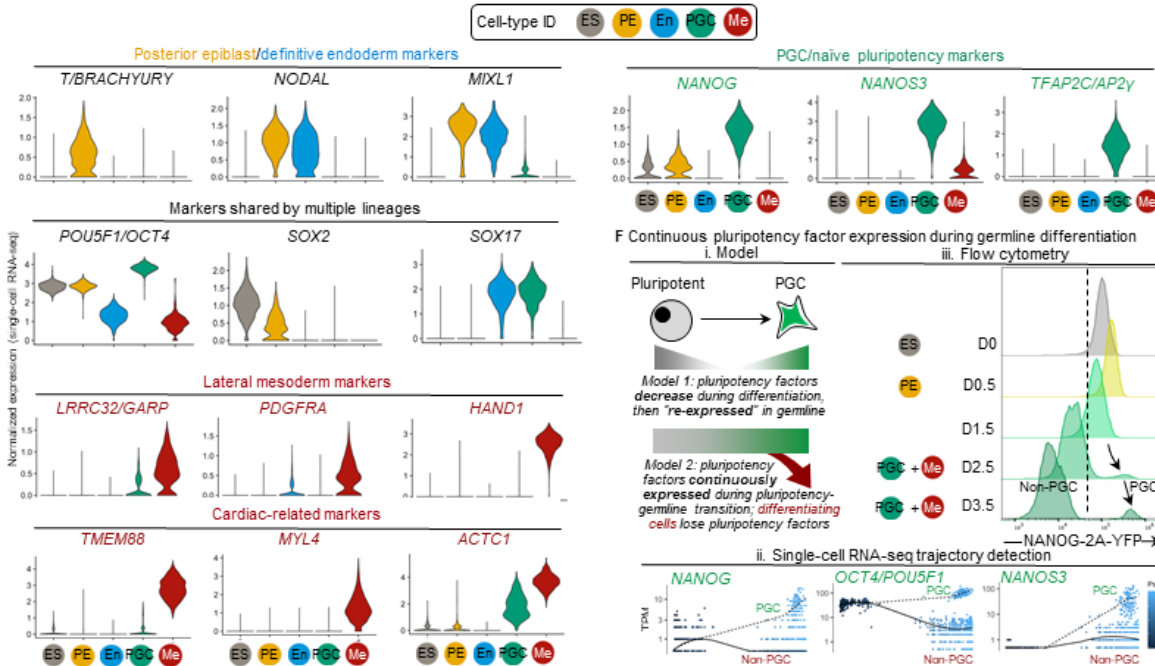
F Continuous pluripotency factor expression during germline differentiation
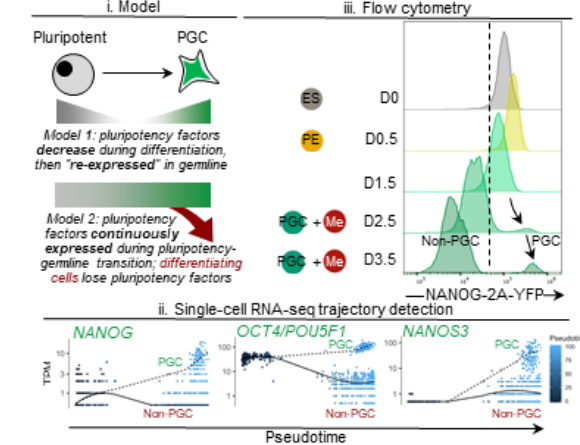
Figure 5

Single cell RNA-sequencing reveals stepwise changes in gene expression, transcriptional trajectories and cellular diversity during hPSC differentiation to PGCLCs a. Schematic of stages profiled for single cell RNA-sequencing (scRNAseq): D0 hPSCs, D0.5 posterior epiblast, D3.5 bulk population, D3.5 FACS-sorted CXCR4+/GARP-/ PDGFRα- PGCLCs and D2 definitive endoderm (left); t-SNE projection of the combined scRNAseq data sets, where single cells are colored by their cluster annotation (right) b. t-SNE projection of hPSC-derived D3.5 bulk population shows that it is heterogeneous and segregates into 2 major clusters: a PGCLC cluster expressing PGC markers (TFAP2C, KLF4, NANOS3) and mesoderm-like cells (non-PGCLCs) expressing mesoderm markers (HAND1, TMEM88, MYL4) c. Immunostaining of hPSC-derived D3.5 bulk population confirms that it is heterogeneous, comprising a mixture of PGCLCs (SOX17+, NANOG+) and non-PGCLCs (HAND1+) (nuclear counterstain: DAPI). Scale bar = 100μm. d. t-SNE projection of scRNAseq data from hPSC-derived FACS-sorted D3.5 CXCR4+/GARP-/PDGFRα- PGCLCs shows that is largely uniform, with the predominant cluster (comprising 97.2% of sorted cells) expressing PGC markers (NANOS3 and TFAP2C) e. Violin plots of scRNAseq data shows expression of posterior epiblast, pluripotency, lateral mesoderm, cardiac, PGC and naive pluripotency markers across the 5 different cell-types (clusters) identified from the combined scRNAseq dataset (comprising merged D0, D0.5 posterior epiblast, D3.5 bulk, D3.5 FACS-sorted PGCLCs and definitive endoderm scRNAseq datasets). f. Different models for pluripotency transcription factor expression during hPSC differentiation into PGCLCs (i, top left); scRNAseq reconstruction of differentiation trajectory of hPSCs progressing into D3.5 PGCLCs and non-PGCLCs, showing pluripotency and PGC markers expression as a function of pseudotime (Qiu et al., 2017) (ii, bottom); flow cytometry analysis of NANOG-2A-YFP knock-in reporter hPSCs differentiating towards D3.5 PGCLCs, showing consistent NANOG expression throughout differentiation (iii, top right)

Fig. 6

A Comparing human fetal PGCs *in vivo* vs. PGCLCs *in vitro*
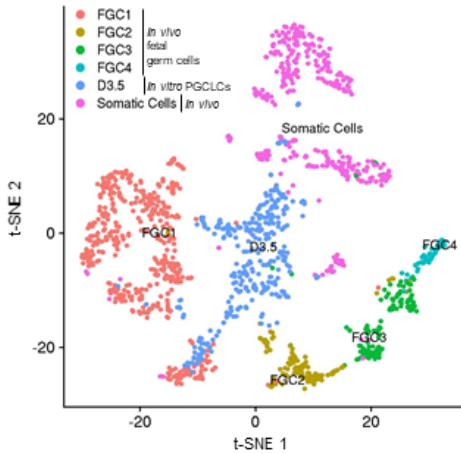
B Comparing human fetal PGCs *in vivo* vs. PGCLCs *in vitro*
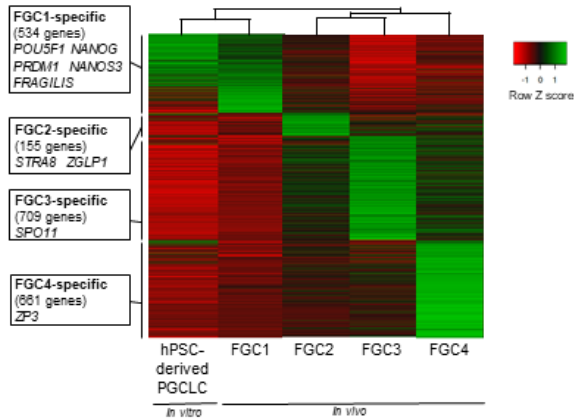
C Pearson correlation between fetal PGCs and PGCLCs
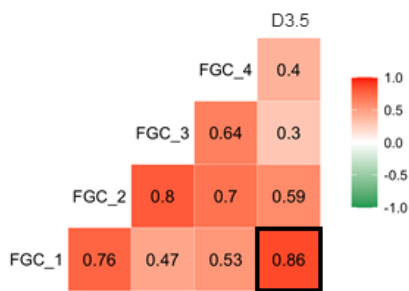
D Gene expression similarities between fetal PGCs and PGCLCs
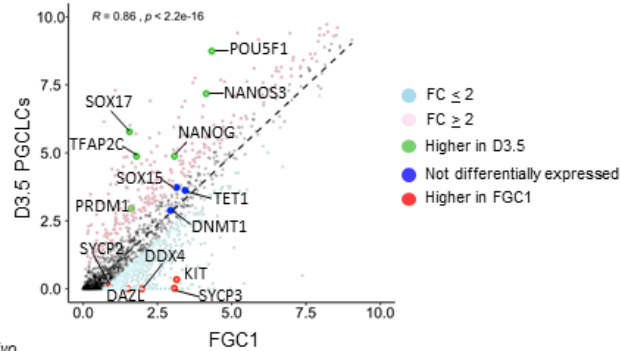
E PGC marker genes expressed by both PGCLCs *in vitro* and PGCs *in vivo*
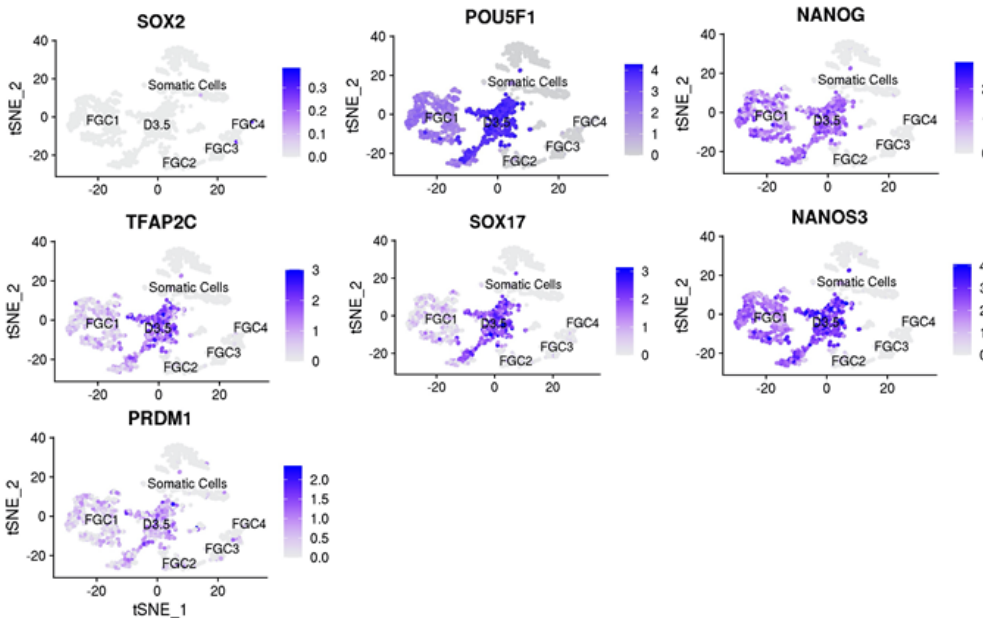
Figure 6

Single cell RNA-sequencing confirms that in vitro-derived PGCLCs resemble human fetal PGCs in vivo at the transcriptome-wide level a. t-SNE plot of scRNAseq profiles of hPSC-derived FACS-sorted CXCR4+ PDGFRA- GARP- D3.5 PGCLCs, together with fetal germ cells (FGCs) and gonadal somatic cells obtained from female human fetuses (Li et al., 2017). FGC1, FGC2, FGC3 and FGC4 represent the four previously-identified transcriptional subsets of FGCs (Li et al., 2017). b. Heatmap of top differentially-expressed

genes among the four subsets of FGCs (FGC1 to FGC4) (Li et al., 2017); expression of these genes in hPSC-derived FACS-sorted CXCR4+ PDGFRA- GARP- D3.5 PGCLCs is also shown. Hierarchical clustering reveals that the hPSC-derived D3.5 PGCLC expression profile clusters with the FGC1 expression profile. Gene expression quantified by scRNAseq. c. Pairwise Pearson correlation analysis between human fetal PGCs (FGC1, FGC2, FGC3, FGC4) and in vitro-derived D3.5 PGCLCs reveals that PGCLCs most closely resemble PGCLCs. Gene expression quantified by scRNAseq. d. Scatterplot demonstrating similarities and differences in gene expression between human fetal PGCs (FGC1) and in vitro-derived D3.5 PGCLCs; each dot represents a single gene. Gene expression quantified by scRNAseq. e. Expression patterns of early PGC markers (as assayed by scRNAseq), depicted on t-SNE plots of in vivo-derived FGCs (Li et al., 2017) and in vitro-derived FACS-sorted CXCR4+ PDGFRA- GARP- D3.5 PGCLCs

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PGCpaperTableDEGsnew.xlsx
- KangetalSuppl.figuresfinal.pptx