Biology of Reproduction, 2017, 97(6), 850–861
doi:10.1093/biolre/iox138
Research Article
Advance Access Publication Date: 30 October 2017

Research Article

Germline competency of human embryonic stem cells depends on eomesodermin†

Di Chen[1], Wanlu Liu[2], Anastasia Lukianchikov[1], Grace V. Hancock[1,2], Jill Zimmerman[1], Matthew G. Lowe[1,2], Rachel Kim[3], Zoran Galic[3,4], Naoko Irie[5,6,7], M. Azim Surani[5,6,7], Steven E. Jacobsen[1,2,3,8,9] and Amander T. Clark[1,2,3,10,*]

[1]Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, California, USA; [2]Molecular Biology Institute, University of California, Los Angeles, California, USA; [3]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, California, USA; [4]Department of Medicine, University of California, Los Angeles, California, USA; [5]Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK; [6]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK; [7]Wellcome Trust-Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, UK; [8]Department of Biological Chemistry, University of California, Los Angeles, California, USA; [9]Howard Hughes Medical Institute, University of California, Los Angeles, California, USA and [10]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, California, USA

*__Correspondence:__ Department of Molecular Cell and Developmental Biology, 615 Charles E Young Drive South, University of California Los Angeles, Los Angeles, CA 90095, USA. E-mail: clarka@ucla.edu

†__Grant Support:__ All experiments, with the exception of human hESC derivation, were funded by R01 HD079546 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) (ATC) and supported by Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research NIH-NCATS UCLA CTSI Grant Number UL1TR0001881. No NIH funds were used for human embryo culture or hESC derivation. Instead, the derivation of hESC lines was funded by the California Institute for Regenerative Medicine, the UCLA Eli, and Edythe Broad Center of Regenerative Medicine and Stem Cell Research (BSCRC). Di Chen is supported by a training grant from the UCLA BSCRC. Wanlu Liu is supported by Philip J. Whitcome fellowship from the UCLA Molecular Biology Institute and a scholarship from the Chinese Scholarship Council. Human fetal tissue was obtained from the Laboratory of Developmental Biology, University of Washington, Seattle, which is supported by NIH Award Number 5R24HD000836 from the NICHD.

Received 23 October 2017; Accepted 27 October 2017

Abstract

In humans, germline competency and the specification of primordial germ cells (PGCs) are thought to occur in a restricted developmental window during early embryogenesis. Despite the importance of specifying the appropriate number of PGCs for human reproduction, the molecular mechanisms governing PGC formation remain largely unexplored. Here, we compared PGC-like cell (PGCLC) differentiation from 18 independently derived human embryonic stem cell (hESC) lines, and discovered that the expression of primitive streak genes were positively associated with hESC germline competency. Furthermore, we show that chemical inhibition of TGFβ and WNT signaling, which are required for primitive streak formation and CRISPR/Cas9 deletion of Eomesodermin (*EOMES*), significantly impacts PGCLC differentiation from hESCs. Taken together, our results suggest that human PGC formation involves signaling and transcriptional programs associated with somatic germ layer induction and expression of EOMES.

Summary Sentence

EOMES induction in the progenitor cell prior to germ cell formation in vitro from hESCs is required for efficient PGC-like cell formation.

Key words: human, EOMES, primordial germ cells, embryonic stem cells.

Introduction

Primordial germ cells (PGCs) are diploid embryonic progenitor cells that ultimately differentiate into haploid gametes. PGCs are specified, maintained, and differentiated in a step-wise manner, and as a result are responsible for the number and quality of adult gametes. Very little is known about the earliest steps in human PGC differentiation in the embryo. Most of our knowledge comes from the mouse [1, 2], and more recently the nonhuman primate cynomolgus (cyno) macaque [3], rhesus macaque [4], and porcine [5]. Analysis of PGCs from cyno and porcine embryos relative to the mouse strongly suggest that the molecular and cellular events in PGC specification are different. Notably, around the time of PGC specification, cyno and porcine embryos develop morphologically as a bilaminar disk, whereas mouse embryos develop as an egg cylinder. In mouse embryos, PR domain zinc finger protein 1 (PRDM1), also called BLIMP1, transcription factor AP2 gamma (TFAP2C), and PR domain zinc finger protein 14 (PRDM14) constitute the critical tripartite transcription factor network responsible for specifying PGC fate from competent epiblast cells in vivo [6–8] and epiblast-like cells in vitro [9, 10]. In contrast, PRDM14 may not be required for human PGC fate [11], and instead, SOX17 has emerged as a new critical regulator in primates [12]. Therefore, the transcriptional network required for PGC specification within the mammalian class appears to have diversified.

Just as the transcriptional network for PGC specification has changed in different mammalian species, the timing of PGC formation in the peri-implantation embryo may have also diverged. In cyno embryos, PGCs are first identified in the peri-implantation embryo prior to primitive streak formation in an extra embryonic cell layer called the amnion [3]. Notably, the amnion is derived from the inner cell mass/epiblast just prior to primitive streak formation. In porcine embryos, PGC specification begins in the proximal pre-streak epiblast [5], whereas in the mouse embryo, PGC precursors are first identified as a small cluster in the proximal epiblast around the time of primitive streak formation [6, 13, 14]. Therefore, in cyno, mouse, and porcine embryos, where PGC specification has been studied in detail, PGC formation is induced from embryonic progenitors at or just prior to formation of the primitive streak and would therefore be exposed to signaling events that establish the primary germlayers and the primitive streak.

One of the earliest markers of primitive streak is the T-box transcription factor BRACHYURY (T), and in mouse embryos, T functions downstream of Wnt signaling to promote PGC specification. [15]. A second T-box transcription factor eomesodermin (EOMES) is also expressed in the proximal epiblast prior to overt primitive streak formation, becoming restricted to the primitive streak during gastrulation [16]. In the mouse, there is no functional evidence for EOMES in the regulation of PGC development. However, using human-induced pluripotent stem cells (hiPSCs) a recent study indicated that EOMES is required for human PGC-like cell (PGCLC) formation following induced reprogramming [17]. It is not known whether EOMES is required for differentiation of human PGCLCs from human embryonic stem cells (hESCs).

Although transcription factors regulating PGC specification may have diverged between mice and humans [12], the major signaling pathways that specify PGCs have not. Most notably, bone morphogenetic protein 4 (BMP4) is essential for mouse PGC formation in vivo [18] as well as human and cyno PGCLC formation in vitro [5, 11, 12, 19, 20]. In the mouse, BMP4 specifies PGCs from the Wnt3-primed posterior epiblast, with T functioning downstream of WNT to regulate PGC fate [15]. In cyno's, T is expressed in the nucleus of nascent PGCs in the amnion, presumably due to Wnt3A signaling [3], and in porcine embryos, T is expressed in nascent PGCs of the preprimitive-streak embryo. Furthermore, porcine PGC numbers diminish following WNT inhibition [5]. These results suggest a conserved role for Wnt and BMP signaling in mouse, porcine, cyno, and human PGC formation in the embryo.

Human PGCLCs induced from multiple hiPSCs have been used to study the germ cell fate determinants [17, 21]. However, much less is known about the competency across different hESCs. In the current study, we discovered a positive correlation between the expression of primitive streak genes and efficiency of human PGCLC induction from 18 independently derived hESC lines. Using CRISPR/Cas9, we provide direct evidence that the transcription factor EOMES is required for efficient induction of human PGCLCs from hESCs, and lends additional support to the hypothesis that human PGCs are induced from cells in the embryo that express EOMES.

Material and methods

Human fetal samples

Human prenatal testes and ovaries were acquired following elected termination and pathological evaluation after University of California, Los Angeles (UCLA)-Institutional Review Board (IRB) review, which deemed the project exempt under 45 CRF 46.102(f). All prenatal gonads were obtained from the University of Washington Birth Defects Research Laboratory (BDRL), under the regulatory oversight of the University of Washington IRB-approved Human Subjects protocol combined with a Certificate of Confidentiality from the Federal Government. BDRL collected the fetal testes and ovaries and shipped them overnight in HBSS with ice pack for immediate processing in Los Angeles. All consented material was donated anonymously and carried no personal identifiers. Developmental age was documented by BDRL as days post fertilization using a combination of prenatal intakes, foot length, Streeter Stages, and crown-rump length. All prenatal gonads documented with birth defect or chromosomal abnormality, were excluded from this study.

hESC culture

All primed hESC lines were cultured on mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) in hESC media, which is composed of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100 μM L-Glutamine (GIBCO, 25030-081), 1× MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 55 μM 2-Mercaptoethanol (GIBCO, 21985-023), 10 ng/mL recombinant

human FGF basic (R&D systems, 233-FB), 1× Penicillin-Streptomycin (GIBCO, 15140-122), and 50 ng/mL primocin (InvivoGen, ant-pm-2) in DMEM/F12 media (GIBCO, 11330-032). All hESC lines were split every 7 days with Collagenase type IV (GIBCO, 17104-019). 4i hESCs were maintained as described before [19]. 4i cells were grown on irradiated MEFs (GlobalStem) in knockout DMEM containing 20% KSR, 2 mM L-glutamine, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol (all GIBCO), 20 ng/mL human LIF (Stem Cell Institute [SCI]), 8 ng/mL bFGF (SCI), 1 ng/mL TGF-β1 (Peprotech), 3 μM CHIR99021 (Miltenyi Biotec), 1 μM PD0325901 (Miltenyi Biotec), 5 μM SB203580 (TOCRIS bioscience), and 5 μM SP600125 (TOCRIS bioscience). 4i hESCs were split every 3 to 5 days using TrypLE Express (GIBCO). An amount of 10 μM of ROCK inhibitor (Y-27632, TOCRIS bioscience) was used for the first 24 h after passage. All hESC lines used in this study are registered with the National Institute of Health Human Embryonic Stem Cell Registry and are available for research use with NIH funds. Specifically, the following hESC lines were used in this study: UCLA1 (46XX), UCLA2 (46XY), UCLA3 (46XX), UCLA4 (46XX), UCLA5 (46XX), UCLA6 (46XY), UCLA7 (47XX+13), UCLA8 (46XX), UCLA9 (46XX), UCLA10 (46XY), UCLA11 (46XY), UCLA12 (46XX), UCLA13 (46XY), UCLA14 (46XX), UCLA15 (46XX), UCLA16 (46XX), UCLA17 (46XX), UCLA18 (46XX). The derivation and basic characterization (Karyotype and teratoma analysis) of UCLA1–6 were previously reported [22]. UCLA8–10, UCLA14, and UCLA16–18 were reported [23].

Derivation and characterization of ESC lines from human embryos

The following UCLA hESC lines UCLA7 (47XX+13), UCLA11 (46XY), UCLA12 (46XX), UCLA13 (46XY), UCLA15 (46XX) were derived from human embryos according to the methods described [22]. All hESC derivations were performed after human subjects' approval from the UCLA-IRB and following Embryonic Stem Cell Research Oversight committee approval. No NIH funds were used for the derivation or initial characterization (karyotype and teratoma) of hESC lines. Teratoma analysis was performed after Institutional approval by the UCLA Office of Animal Research Oversight. Teratomas were created by injecting of collagenase digested clumps of hESCs into the testicles of male scid/beige C.B.17-Prkdc(scid)Lyst(bg) mice. Prior to injection, hESCs were resuspended in ice-cold matrigel (Corning, 354277), with 3× wells (from a 6-well plate) of colonies injected per testis. Karyotype analysis was conducted by Cell Line Genetics.

Induction of human PGCLCs though incipient mesoderm-like cells from primed hESCs

Human PGCLCs were induced from primed hESCs as described in [19] with some modifications. Day 7 hESCs were dissociated into single cells with 0.05% Trypsin-EDTA (Gibco, 25300-054) and plated onto Human Plasma Fibronectin (Invitrogen, 33016-015)-coated 12-well plate at the density of 200 000 cells/well in 2 mL/well of incipient mesoderm-like cell (iMeLC) media, which is composed of 15% KSR, 1× NEAA, 0.1 mM 2-Mercaptoethanol, 1× Penicillin-Streptomycin-Glutamine (Gibco, 10378-016), 1 mM sodium pyruvate (Gibco, 11360-070), 50 ng/mL Activin A (Peprotech, AF-120-14E), 3 μM CHIR99021 (Stemgent, 04-0004), 10 μM of ROCKi (Y27632, Stemgent, 04-0012-10), and 50 ng/mL primocin in Glasgow MEM (GMEM) (Gibco, 11710-035). iMeLCs were dissociated into single cells with 0.05% Trypsin-EDTA after 24 h of incubation unless otherwise mentioned and plated into ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at the density of 3000 cells/well in 200 μL/well of PGCLC media, which is composed of 15% KSR, 1× NEAA, 0.1 mM 2-Mercaptoethanol, 1× Penicillin-Streptomycin-Glutamine (Gibco, 10378-016), 1 mM sodium pyruvate (Gibco, 11360-070), 10 ng/mL human LIF (Millipore, LIF1005), 200 ng/mL human BMP4 (R&D systems, 314-BP), 50 ng/mL human EGF (R&D systems, 236-EG), 10 μM of ROCKi (Y27632, Stemgent, 04-0012-10), and 50 ng/mL primocin in GMEM (Gibco, 11710-035). An amount of 100 ng/mL stem cell factor (SCF; PEPROTECH, 250-03), 10 μM SB431542 (Stemgent, 04-0010-10), and 500 ng/μL Dickkopf Wnt Signaling Pathway Inhibitor 1 (DKK1) (R&D systmes, 5439-DK) was added in PGCLC media for some experiments.

Induction of human PGCLCs from 4i hESCs

4i WIS2 cells were dissociated with TrypLE and plated to ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at the density of 2000–4000 cells/well in 200 μL PGCLC media, which is composed of 15% KSR, 0.1 mM NEAA, 0.1 mM 2-mercaptoethanol, 100 U/mL Penicillin–0.1 mg/mL Streptomycin, 2 mM L-Glutamine, 1 mM Sodium pyruvate, 500 ng/mL BMP4 (R&D Systems) or BMP2 (SCI), 1 μg/mL human LIF (SCI), 100 ng/mL SCF (R&D Systems), 50 ng/mL EGF (R&D Systems), and 10 μM ROCK inhibitor in GMEM (Gibco, 11710-035).

Flow cytometry and fluorescence activated cell sorting

Human prenatal gonads or day 4 aggregates were dissociated with 0.25% trypsin (Gibco, 25200-056) for 5 min or 0.05% Trypsin-EDTA (Gibco, 25300-054) for 10 min at 37°C. The dissociated cells were stained with conjugated antibodies, washed with fluorescence activated cell sorting (FACS) buffer (1% BSA in PBS), and resuspended in FACS buffer accompanying with 7-AAD (BD Pharmingen, 559925). The single cell suspension was analyzed or sorted. The conjugated antibodies used in this study are ITGA6 conjugated with BV421 (BioLegend, 313624), EPCAM conjugated with 488 (BioLegend, 324210), EPCAM conjugated with APC (BioLegend, 324208), tissue nonspecific alkaline phosphatase (TNAP) conjugated with PE (BD Pharmingen, 561433), and cKIT conjugated with APC (BD Pharmingen, 550412).

Real-time quantitative polymerase chain reaction

Sorted cells or cell pellets were lysed in 350 μL of RLT buffer (QIAGEN) and RNA was extracted using RNeasy micro kit (QIAGEN, 74004). Complementary DNA was synthesized using SuperScript II Reverse Transcriptase (Invitrogen, 18064-014). Real-time quantitative polymerase chain reaction (PCR) was performed using TaqMan Universal PCR Master Mix (Applied Biosystems, 4304437) and the expression level of genes-of-interest were normalized to the expression of housekeeping gene GAPDH. The Taqman probes used in this study include GAPDH (Applied Biosystems, Hs99999905_m1), NANOS3 (Applied Biosystems, Hs00928455_s1), PRDM1 (Applied Biosystems, hs01068508_m1), TFAP2C (Applied Biosystems, Hs00231476_m1), SOX17 (Applied Biosystems, Hs00751752_s1), cKIT (Applied Biosystems, hs00174029_m1), DAZL (Applied Biosystems, hs00154706_m1), DDX4 (Applied Biosystems, Hs00251859_m1), SOX9 (Applied Biosystems, Hs01001343_g1), AMH (Applied Biosystems, Hs01006984_g1), DND1 (Applied Biosystems, Hs00832091_s1), T (Applied Biosystems, Hs00610080_m1), CER1 (Applied Biosystems,

Hs00193796_m1), MIXL1 (Applied Biosystems, Hs00430824_g1), WNT3 (Applied Biosystems, Hs00902257_m1), EOMES (Applied Biosystems, Hs00172872_m1), GSC (Applied Biosystems, Hs00906630_g1), FOXA2 (Applied Biosystems, Hs00232764_m1), POU3F1 (Applied Biosystems, Hs00538614_s1), TCF15 (Applied Biosystems, Hs00231821_m1). For human gonad samples, two technical replicates of real-time quantitative PCR were performed. For hESCs, iMeLCs, and PGCLCs, two to three independent experiments were performed.

Generation of *EOMES* mutant hESC lines

A pair of guide RNA (gRNA) targeting EOMES was designed using crispr.mit.edu, and the corresponding gDNA sequence was cloned into px330 vector [24]. An amount of 4 μg of gRNA pair or 2 μg of pMax-GFP was electroporated into 800 000 UCLA1 cells using P3 Primary Cell 4D-Nucleofector X Kit according to the manufacturer's instructions (Lonza, V4XP-3024). Twenty-four hours after nucleofection, cells were dissociated with Accutase (ThermoFisher Scientific, A1110501) and seeded in low density. A total of 96 individual colonies were picked after 9 days and expanded. Lines were screened for the presence of shorter bands due to deletion. To determine the precise mutations, PCR product from the targeted allele was cloned using Topo-TA cloning (Thermo-Fisher) and analyzed by Sanger sequencing. Two mutant lines were chosen and subcloned before experiments. The gRNA sequences used to target *EOMES* are 5'- GCGGTGTACAGCCGTAACAT and 5'-GTTATCTACACCGAAAGTGC. Karyotyping by Cell Line Genetics was performed before experimentation. GFP-labeled control and EOMES mutant hESCs were made by lentivirus-based transfection of UbiC-GFP-IRES-Puromycin and maintained as stable cell lines with puromycin (1 μg/mL) selection.

Immunofluorescence

Immunostaining paraffin sections or aggregates in whole mount were described previously [25, 26]. For cells cultured on chamber slides, samples were fixed in 4% paraformaldehyde in PBS for 10 min and washed with PBS containing 0.1% Tween 20 and permeabilized with PBS containing Triton X for 10 min. Samples were blocked with 10% donkey serum for 30 min before antibody incubation. The primary antibodies used for immunofluorescence in this study include rabbit-anti-EPCAM (Abcam, ab71916, 1:50), goat-anti-VASA (R&D Systems, AF2030, 1:20), rabbit-anti-cKIT (DAKO, A4502, 1:100), goat-anti-OCT4 (Santa Cruz Biotechnology, sc-8628x, 1:100), rabbit-anti-PRDM1 (Cell Signaling Technology, 9115, 1:100), mouse-anti-PRDM1 (R&D Systems, MAB36081SP, 1:100), rabbit-anti-TFAP2C (Santa Cruz Biotechnology, sc-8977, 1:100), mouse-anti-TFAP2C (Santa Cruz Biotechnology, sc12762, 1:100), rat-anti-ITGA6 (Santa Cruz Biotechnology, sc-80554, 1:100), goat-anti-T (R&D Systems, AF2085, 1:100), goat-anti-SOX17 (Neuromics, GT15094, 1:100), rabbit-anti-EOMES (Abcam, ab23345, 1:100), rabbit-anti- β-CATENIN (Cell Signaling Technology, 9582, 1:100), rabbit-anti-pSMAD2/3 (Cell Signaling Technology, 8828, 1:100). The secondary antibodies used in this study are all from Jackson ImmunoResearch Laboratories including donkey-anti-rabbit-488, donkey-anti-mouse-488, donkey-anti-goat-488, donkey-anti-rat-488, donkey-anti-rabbit-594, donkey-anti-mouse-594, and donkey-anti-goat-594. DAPI is counterstained to indicate nuclei.

RNA-sequencing

Sorted cells or cell pellets were lysed in 350 μL of RLT buffer (QIAGEN), and total RNA was extracted with RNeasy micro kit (QIAGEN, 74004). Total RNA was reverse transcribed and cDNA was amplified using Ovation RNA-Seq System V2 (Nugen, 7102-32) according to manufacturer's instructions. Amplified cDNA was fragmented into ~200 bp by Covaris S220 Focused-ultrasonicators. RNA-sequencing (RNA-seq) libraries were generated using Ovation Rapid Library Systems (Nugen, 0319-32 for index 1-8 and 0320-32 for index 9–16) and quantified by KAPA library quantification kit (Illumina, KK4824). Libraries were subjected to single-end 50 bp sequencing on HiSeq 2000 or HiSeq 2500 sequencer with 4–6 indexed libraries per lane.

RNAseq analysis

Analysis of individual gene expression

Raw reads in qseq format obtained from sequencer were first converted to fastq format with customized perl script. Reads quality were controlled with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). High-quality reads were then aligned to hg19 reference genome using Tophat [27] (v 2.0.13) by using "-no-coverage-search" option, allowing up to two mismatches and only keeping reads that mapped to one location. Basically, reads were first mapped to hg19 gene annotation with known splice junction. When reads did not map to the annotated genes, the reads were mapped to hg19 genome. Number of reads mapping to genes were calculated by HTseq [28] (v 0.5.4) with default parameters. Expression levels were determined by RPKM (reads per kilobase of exons per million aligned reads) in R using customized scripts. For RNAseq of published datasets GSE60138 [12], GSM1643143 [19], raw reads were obtained from GEO and then processed exactly the same as described above.

Hierarchical clustering of RNAseq

Raw read counts for each gene obtained from HTseq were preprocessed with DESeq R package [28]. To account for heteroscedasticity between samples, variance stabilizing transformation was first applied to all genes with DEseq. Samples were then hierarchical clustered (hclust function) based on their Euclidian distances (dist function) in R using customized scripts.

Principal component analysis

For principal component analysis (PCA), RPKM for each sample was first calculated. Variance of each genes across samples were then calculated (rowVars function in R). PCA analysis (prcomp function in R) was performed on genes with the top 500 variation across samples. PCA were then plotted with ggplot2 package in R (http://ggplot2.org).

Correlation analysis between samples

Correlation between ITGA6/EPCAM and TNAP/cKIT was calculated and plotted in R. Firstly, RPKM of each ITGA6/EPCAM and TNAP/cKIT RNAseq replicates were calculated. Pearson correlations were then obtained with cor function. Scatter plot was plotted with ggplot2 with linear regression line. Human PGC-enriched genes compared to H9 hESCs were defined by Irie et al. [12] (only genes with at least four fold enrichment in hPGC compared to H9 hESCs were selected). The RPKM of those selected genes were extracted, and heat map was plotted with log2(RPKM+1) values in R using pheatmap package.

Differential gene expression calling

R DESeq package was used to normalize counts per RefSeq transcripts to evaluate differential expression. For comparison between hESC, iMeLC, mutant hESC, and mutant iMeLC, highly upregulated genes of each sample, with mean log2(fold change) >1 and adjusted P value < 0.05 were selected and plotted as Venn diagram using VennDiagram package in R. Scatter plot of log2(RPKM+1) was plotted in R with ggplot2 package. Log2(Fold change) were colored with different scale.

Analysis of gene expression level and PGCLC induction efficiency

In order to obtain iMeLC-specific genes coexpressing with PGCLC induction efficiency, Spearman correlation of each gene's expression level in iMeLC was calculated with PGCLC induction efficiency. Genes in hESC with RPKM more than 2 were filtered out, while genes with at least 2 RPKM in iMeLC were kept. Gene's expression levels with at least 0.45 correlations to PGCLC induction efficiency were kept. Heat map of log2(RPKM+1) was plotted in R as described above.

Results

ITGA6 and EPCAM can be used to isolate human PGCs from embryonic ovaries but not embryonic testes

Recently, in vitro human PGCLCs were isolated using INTEGRINα6 (ITGA6) and EPCAM following hiPSC differentiation [19]; however, whether ITGA6/EPCAM can be used to isolate in vivo PGCs from human embryos has not been shown. To address this, we examined cells from a pair of human embryonic ovaries at 72 day postfertilization by staining with four antibodies that detect ITGA6, EPCAM, TNAP, and cKIT. cKIT and TNAP were used as a positive control, given that these surface proteins can be used to sort human PGCs from the prenatal embryonic ovary and testis [25, 29]. The labeled cells were divided in two and sorted by FACS gating on either ITGA6/EPCAM or TNAP/cKIT (Figure 1A). We discovered that the percentage of ITGA6/EPCAM double positive cells and TNAP/cKIT double positive cells in the embryonic ovary was comparable (Figure 1A). Furthermore, the ITGA6/EPCAM double positive population was also double positive for TNAP/cKIT. Conversely, the TNAP/cKIT double positive cells were also double positive for ITGA6/EPCAM (Figure 1A). Using real-time PCR, we show that both populations expressed PGC genes at equivalent levels (Figure 1B). We repeated this experiment with a pair of human embryonic ovaries at 94 day, and found a similar result (Figure 1A and B). Therefore, these observations indicate that ITGA6/EPCAM can be used to isolate TNAP/cKIT positive germ cells from the human embryonic ovary and vice versa.

Next, we performed the same analysis using embryonic testes. Unlike the ovary, we discovered that most ITGA6/EPCAM double positive cells are negative for TNAP/cKIT (Figure 1C), and the ITGA6/EPCAM double positive cells have reduced levels of PGC genes relative to the TNAP/cKIT double positive population (Figure 1D). Given that ITGA6 and EPCAM are epithelial markers [30, 31], and PGCs in the embryonic testis are known to be enclosed within epithelial cords [25], we reasoned that ITGA6 and EPCAM double positive cells are a mixture of both germ cells and somatic cells in prenatal testis. To test this, we performed immunofluorescence and real-time PCR and show that ITGA6 and EPCAM are expressed by both PGCs and epithelial Sertoli cells (Figure S1A and B), and the ITGA6/EPCAM double positive population is also enriched in Sertoli cell genes *SOX9* and *AMH* (Figure 1D). Therefore, using prenatal tissues containing PGCs, TNAP/cKIT can be used to sort PGCs from both the embryonic ovary and testis, whereas ITGA6/EPCAM can only be used to sort PGCs from the embryonic ovary.

Next, we performed RNA-seq on sorted ITGA6/EPCAM female PGCs at 89 and 103 day (Figure 1E and F), and compared this to a third sample involving a pair of ovaries collected at 89 d, which were pooled and then stained and sorted separately for ITGA6/EPCAM or TNAP/cKIT (Figure 1E–G, marked by asterisk). We also included a testis at 59 day, which was sorted only using TNAP/cKIT to specifically isolate the male PGCs. Using unsupervised hierarchical clustering (UHC), the sorted PGC samples clustered together forming a distinct group relative to an undifferentiated female hESC line (UCLA1), and a male hESC line (UCLA2) (Figure 1E). Evaluation of candidate genes revealed that all putative PGC samples expressed the early and late stage PGC genes and have acquired a naïve pluripotent signature (Figure S1C). Moreover, the ITGA6/EPCAM and TNAP/cKIT PGCs sorted from the same pair of 89 day embryonic ovaries clustered the closest compared to ITGA6/EPCAM populations sorted from different ovaries (Figure 1E). This result further supports the conclusion that ITGA6/EPCAM and TNAP/cKIT are expressed on the same population of PGCs. To better display the similarities and differences between samples, we performed PCA, and found that the PGCs are well separated from undifferentiated hESCs along the PC1 axis, while the 59-day male TNAP/cKIT PGCs are separated from the other gonadal PGCs along the PC2 axis (Figure 1F). This separation is anticipated to be caused by the 4-week difference in age between male and female samples in this study, as well as gender [32]. Last, we examined the differentially expressed genes (DEGs) between ITGA6/EPCAM and TNAP/cKIT positive cells sorted from the same ovary pair and found very few DEGs as the two transcriptomes are highly correlated with linear regression R squared = 0.98 (Figure 1G). Therefore, we propose that ITGA6/EPCAM and TNAP/cKIT are coexpressed on the same cell population in the embryonic ovary.

Germline competency is associated with the induction of genes required for primitive streak formation

Prior to examining PGCLC differentiation across 18 independently derived UCLA hESC lines with ITGA6/EPCAM/TNAP/cKIT, we first piloted the sorting strategy on the male hESC line UCLA2. We discovered that all ITGA6/EPCAM double positive putative PGCLCs at day 4 of differentiation were positive for TNAP; however, cKIT was not detected on the PGCLC population (Figure S2A). In mice, cKIT is critical for PGC survival, migration, and proliferation [33–35]. In humans, cKIT is present on PGCs from at least week 4 to week 20 of prenatal life postfertilization [25, 29, 32, 36]. Furthermore, analysis of cyno PGCs during gastrulation reveals that *cKIT* RNA is expressed by PGCs at the time of specification [3]. Thus, we reasoned that cKIT is either not translated in early stage human PGCs, or alternatively the conditions for PGCLC induction must be optimized to enable cKIT expression on the cell surface. Given that cKIT is subject to ligand induced endocytosis [37], we removed the ligand SCF from the PGCLC media, and this resulted in ITGA6/EPCAM/TNAP/cKIT positive PGCLCs (Figure S2B and D). To determine if excluding SCF affects germ cell identity, we performed real-time PCR and RNA-Seq of the putative ITGA6/EPCAM positive PGCLCs differentiated with or without SCF (Figure S2C and E). To determine the molecular similarity of PGCLCs generated in primed conditions on MEFs, to those generated from hESCs cultured in 4i on MEFs [12], we

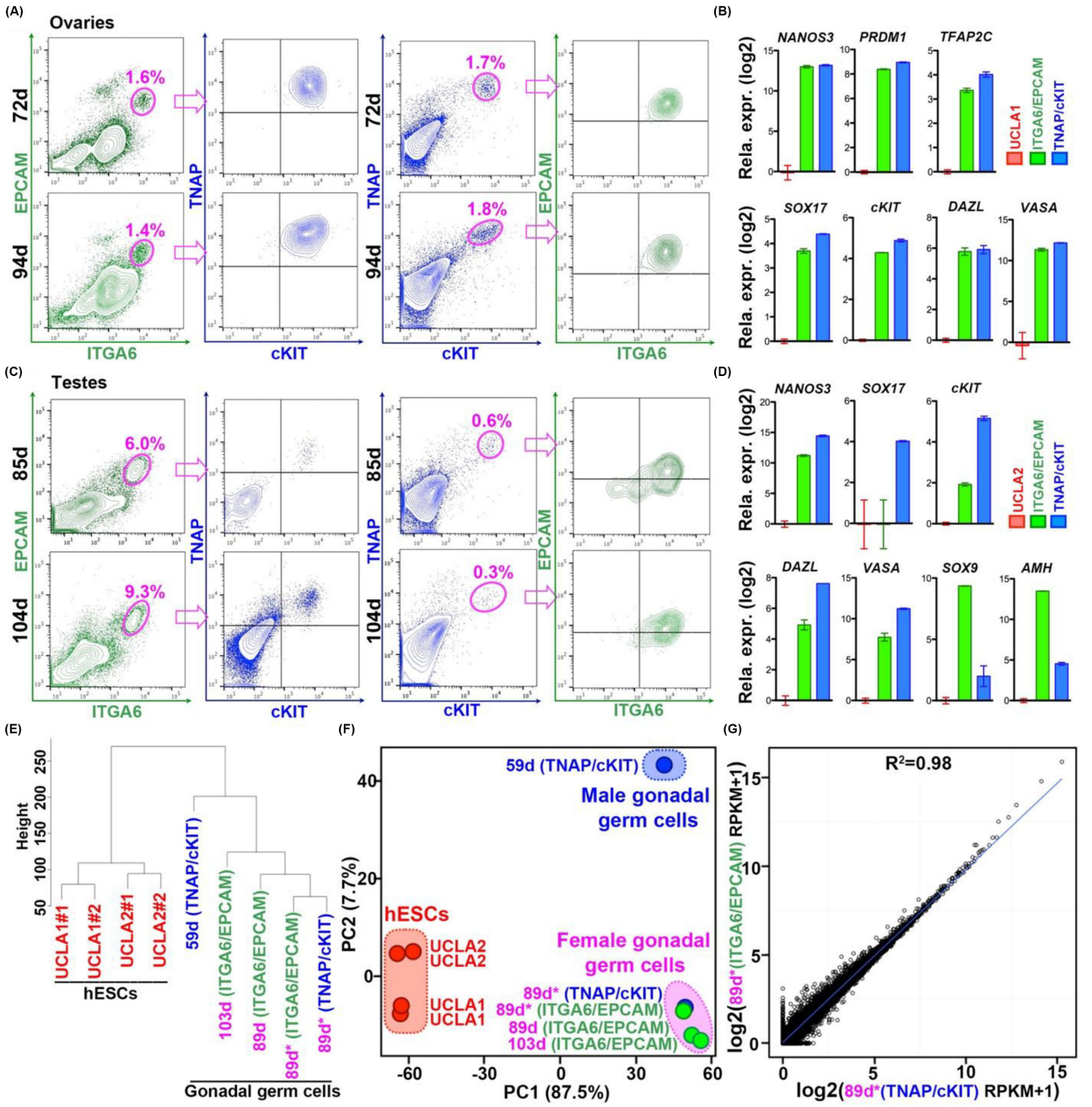

Figure 1. Analysis of ITGA6/EPCAM and TNAP/cKIT populations in human prenatal gonads. (A) Flow cytometry of prenatal ovaries at day (d) 72 and 94 postfertilization stained with antibodies that recognize ITGA6, EPCAM, TNAP, and cKIT. (B) Gene expression of the sorted populations from (A, 72d) by real-time PCR. Expression is normalized to the GAPDH. Fold change is calculated relative to expression levels of each gene in female hESC line UCLA1 (passage 17 (p17) and p18), which was given a value of 1.0. (C) Flow cytometry of prenatal testis at day 85 and 104 postfertilization stained with ITGA6, EPCAM, TNAP, and cKIT. (D) Gene expression of the sorted populations from (C, 85d) by real-time PCR. For each gene examined, its expression is normalized to the GAPDH. Fold change is calculated relative to expression levels of each gene in male hESC line UCLA2 (p11 and p12), which was given a value of 1.0. (E) Unsupervised hierarchical clustering (UHC) of transcriptomes of female hESC line UCLA1 (two biological replicates, p14 and p15), male hESC line UCLA2 (two biological replicates, p13 and p14), TNAP/cKIT positive germ cells from embryonic day 59 testes, ITGA6/EPCAM positive cells from 89d, 103d, and another 89d embryonic ovary. TNAP/cKIT positive cells from 89d ovaries. Asterisk in E–G indicates that the two RNA-seq libraries were made from the same pair of ovaries but sorted with different surface markers. (F) PCA of transcriptomes shown in E. (G) Scatter plot of two transcriptomes made from the same pair of ovaries but sorted with ITGA6/EPCAM (x-axis) and TNAP/cKIT (y-axis).

differentiated the WIS2 hESC line for 4 days according to the methods of Irie et al. [12] and sorted the TNAP/NANOS3-mCherry PGCLC population. RNA-seq analysis showed that PGCLCs generated from either the primed or 4i cultured hESCs clustered together, and were distinct from the undifferentiated hESCs (Figure S2E and F), indicating that the starting culture media (4i on MEFs versus primed media on MEFs) ultimately yielded PGCLCs with similar transcriptional identities.

Using ITGA6/EPCAM as a sorting approach to analyze PGCLCs, we next examined PGCLC competency of 18 hESC lines, all derived at UCLA from 18 single frozen embryos (Figure 2A). All hESC lines were capable of teratoma formation when injected into severe combined immunocompromized (SCID) beige mice (UCLA1–6 [22]; UCLA8–10, UCLA14, UCLA16–18 [23]; and this study UCLA7, UCLA11–13, and UCLA15: Figure S3A–C). Seventeen out of eighteen hESC lines are karyotypically normal, while UCLA7 has a duplication of chromosome 13 in 100% of cells karyotyped (Figure S3B). We included this cell line in the analysis to determine whether aneuploidy may be associated with alterations in PGCLC potential. It has been previously reported that PGCLC induction efficiency is variable between experiments [12, 19]. In order to minimize variability, we induced PGCLCs from 18 hESC lines simultaneously and repeated this experiment four times to determine the average PGCLC induction efficiency for all 18 lines. All 18 independently derived hESC lines were germline competent. However, UCLA6 consistently exhibited the highest germline potential generating on average 35% PGCLCs at day 4 of aggregate formation, whereas UCLA9 had the lowest germline potential generating on average less than 1% PGCLCs at day 4 (Figure 2A; Figure S3D). The aneuploid UCLA7 female hESC line generated a comparable percentage of PGCLCs to the other female hESC lines in the data set. Meanwhile, we found that male hESC lines on average were more competent for PGCLC induction than female (Figure 2B).

In a recent study using nine hiPSC lines, it was determined that primitive streak genes were associated with PGCLC competency [21]. To determine whether this was also the case for hESCs, we performed RNA-seq on the 18 hESC lines, and the corresponding iMeLCs in biological duplicate. We used the following parameters to define genes whose expression levels in iMeLCs are positively correlated with PGCLC induction efficiency. First, we chose genes that were expressed at low levels or not expressed in hESCs by setting the RPKM value < 2. Second, we used the RPKM value > 2 in at least one of the 18 iMeLC samples to select for genes that are upregulated in at least one of the 18 hESC lines induced to become iMeLCs. Third, we correlated gene expression in iMeLCs with PGCLC induction efficiency, and set the cutoff as >0.45. Based on these parameters, we found 78 genes upregulated from hESCs to iMeLCs that were also positively correlated with resulting PGCLC induction efficiency in the aggregates (Figure 2C). Using gene ontogeny (GO) term analysis of these 78 genes by WebGestalt [38, 39], we discovered that the top term was "formation of primary germ layer" (GO: 0001704), which included seven genes: *EOMES, T, GATA6*, *MIXL1*, *GATA4*, *WLS*, and *LHX1*. These genes are associated with primitive streak formation that are known to be induced by NODAL/ACTIVIN A (TGFβ) and WNT [40].

Signaling pathways that promote primitive streak formation are also required for PGCLC induction

Next, we confirmed the relationship between TGFβ and WNT signaling pathways, and the expression of *T* and *EOMES* in iMeLCs. We performed immunofluorescence and identified nuclear accumulation of phosphorylated SMAD2/3 (pSMAD2/3) (Figure 3A) and β-CATENIN (Figure 3B) in both iMeLCs and hESCs, indicating that these cells are capable of responding to both signaling pathways. We then confirmed expression of T (Figure 3C) and EOMES (Figure 3D) protein by immunofluorescence, and as predicted from the RNA-Seq, T was induced in iMeLCs (Figure 3C), and EOMES protein was expressed in both hESCs and in iMeLCs (Figure 3D).

The correlation of germline competency with TGFβ and WNT signaling, as well as *T* and *EOMES* induction promoted us to test if these molecules are critical for PGCLC formation. To block TGFβ signaling we added SB431542 [41], which inhibits the TGFβ type I receptors ALK5, ALK4, and ALK7, and phosphorylation of SMAD2 and SMAD3. This inhibitor does not block the ALKs or SMADs downstream of BMP4. To block WNT receptor binding, we added DKK1 [42] (Figure 3E). We discovered that the addition of these two molecules prevented the induction of *T* and *EOMES* in iMeLCs despite the presence of ACTIVIN A and GSK3βi in the iMeLC media (Figure 3F). In order to determine if *T* and *EOMES* expression is dependent on TGFβ or WNT, or both, we evaluated *T* and *EOMES* expression in iMeLCs in the presence of either SB431542 or DKK1. We found that *T* and *EOMES* were repressed in the presence of either SB431512 or DKK1 (Figure S4A and B), suggesting the expression of *T* and *EOMES* in iMeLCs requires both TGFβ and WNT signaling. To evaluate the effect on germline competency, we induced PGCLCs in media with or without SB431542 and DKK1 (Figure 3E). We discovered that the PGCLC population was completely abolished when SB431542 and DKK1 were included in the media, despite the presence of exogenous BMP4 and other cytokines known to induce PGCLC fate (Figure 3G). These results suggest that the ability to respond to TGFβ and WNT signaling and potentially the induction of genes associated with primitive streak formation such as *T* and *EOMES* are required for germline induction.

EOMES is required for PGCLC induction

To provide direct evidence for EOMES in PGCLC induction, we used CRISPR/Cas9 to mutate *EOMES* in the UCLA1 hESC line (Figure S4C). We used two independent *EOMES* mutant clones for analysis. By differentiating the *EOMES* mutant and control hESC lines in parallel, we found that the ITGA6/EPCAM double positive PGCLC population was dramatically reduced in the *EOMES* mutant clones (Figure 4A–B). To confirm this result, we also performed immunofluorescence on aggregates and detected PGCLCs by triple staining for TFAP2C, SOX17, and PRDM1. In wild-type control cells, we detected clusters of triple TFAP2C, SOX17, and PRDM1 human PGCLCs (Figure 4C, white dot outlined). However, in the *EOMES* mutant lines, we identified very few triple positive PGCLCs (Figure 4C, white arrowheads). Therefore, our results demonstrate that EOMES is required for PGCLC formation in human, most likely downstream of WNT and TGFβ.

In cyno embryos, nascent PGCs in the amnion are negative for *EOMES* [3]. Similarly, our RNA-seq data show that *EOMES* is not expressed in prenatal PGCs or PGCLCs. Therefore, we hypothesize that EOMES acts either noncell autonomously in the niche to support PGCLC formation, or alternatively EOMES acts cell autonomously in iMeLCs to prime PGC fate. To test this, we made stable GFP lines of control and *EOMES* mutant hESCs. We made iMeLCs and generated PGCLC aggregates containing a mix of GFP and unlabeled cells in a 1:1 ratio. The iMeLC mixtures were composed of GFP-labeled control cells with unlabeled controls (Figure 4D) or

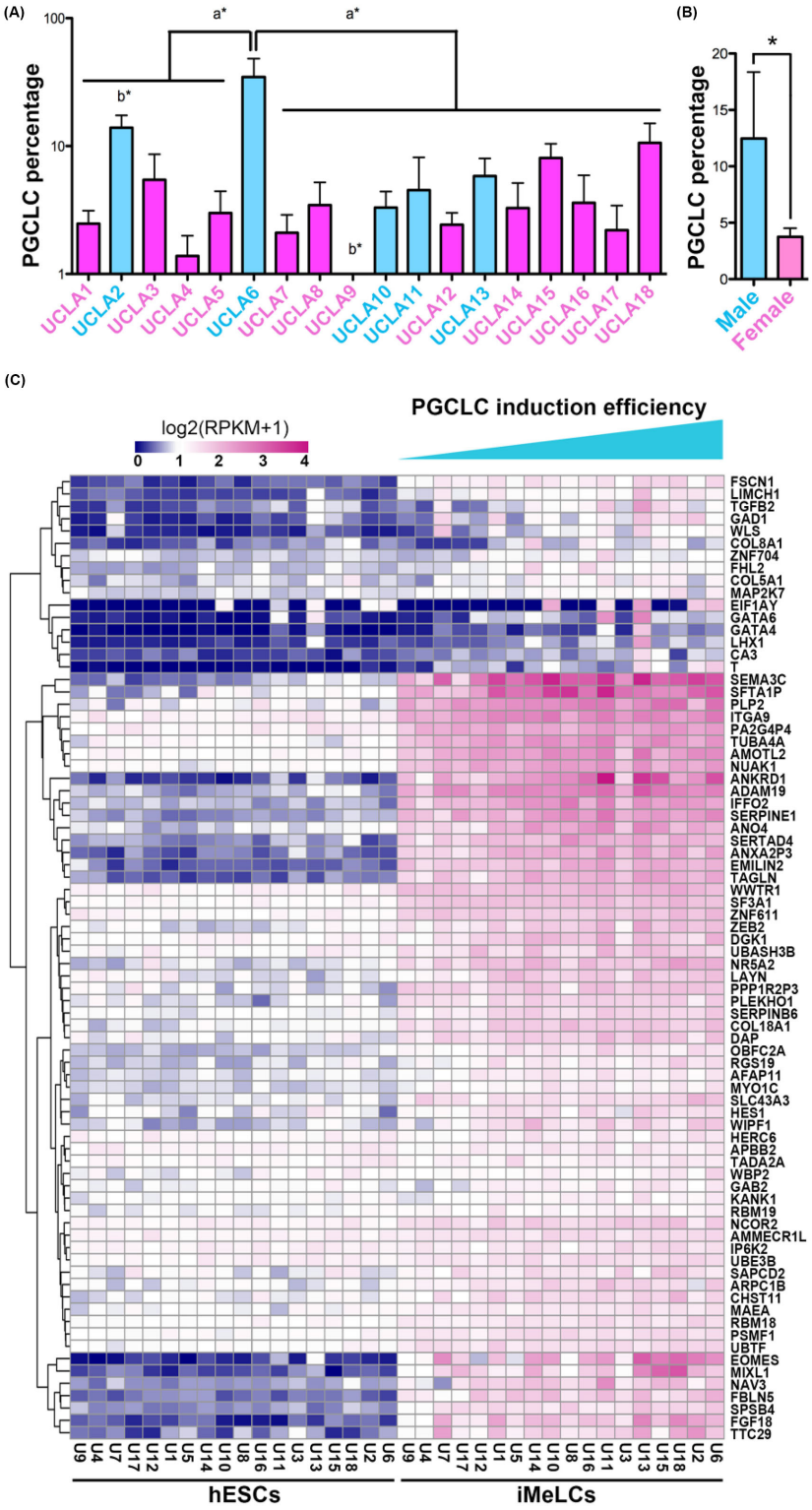

Figure 2. Germline competency varies between independent hESC lines. (A) Average PGCLC induction efficiency at day 4 in aggregates generated from 18 hESC lines (passage numbers ranging from p10 to p22, see experimental procedures for details). Blue represents male and pink represents female. "a*" indicates the significant difference between UCLA6 and all other cell lines and "b*" indicates the significant difference between UCLA2 and UCLA9 (tested by ANOVA, $P < 0.0001$). PGCLCs were identified as ITGA6/EPCAM double positive cells. (B) Comparison of day 4 PGCLC induction efficiency from male (blue) and female (pink) hESC lines. "*" indicates the difference between male and female (t-test, $P < 0.05$). (C) Heat map showing the expression of genes in iMeLCs that positively correlated with PGCLC induction efficiency. Genes are selected as maximal expression <= 2 (RPKM) in hESCs, maximal expression >= 2 (RPKM) in iMeLCs, and correlation coefficient >0.45.

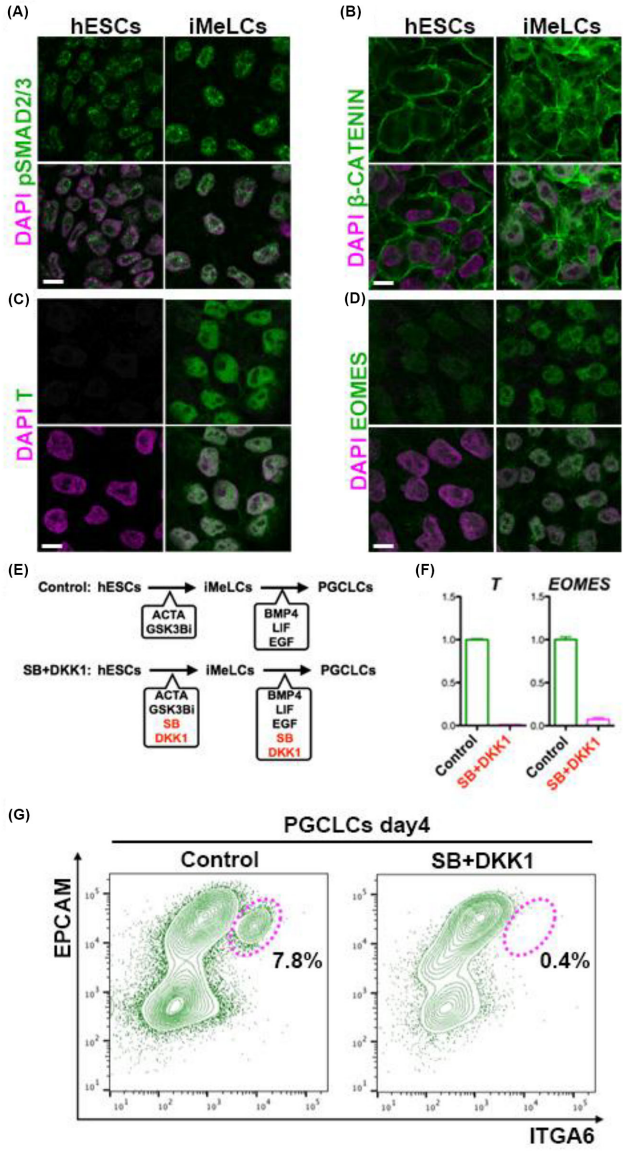

Figure 3. TGFβ and WNT signaling are required for PGCLC induction from hESCs. (A–D) Expression of pSMAD2/3 (A), β-CATENIN (B), T (C), and EOMES (D) in UCLA1 hESCs and iMeLCs. Scale bar: 10 μm. (E) Schematic illustration of PGCLC induction with or without SB431542 (SB) and DKK1. (F) Real-time PCR for *T* and *EOMES* expression at iMeLCs in the presence of SB431542 and DKK1. (G) Flow cytometry showing loss of PGCLC competency in media containing SB431542 and DKK1.

GFP-labeled *EOMES* mutant cells with unlabeled controls (Figure 4E). In the controls, about 45% of the total cells were positive for GFP, and about 35% of the PGCLCs were positive. Therefore, control hESCs with or without GFP both contribute to PGCLC induction (Figure 4D). In contrast, only about 8% of the PGCLCs were induced from GFP-labeled *EOMES* mutant cells, whereas the total cells were composed of about 48% GFP-labeled EOMES mutant cells (Figure 4E). This suggests that EOMES is required cell autonomously to prime PGC fate.

Discussion

In the current study, we prove that PGCLC competency is an inherent property of hESC lines. Together with previous work [11, 12, 19], our study proves that the generation of human PGCLCs in vitro through directed differentiation is not a stochastic event restricted to a small number of hESC or hiPSC lines, but rather a competency that extends to the majority of human pluripotent stem cell lines in vitro regardless of whether they originated from the inner cell mass or through induced reprogramming. We also show that heterogeneity in PGCLC differentiation can provide a unique opportunity to discover conserved and new genes that regulate PGC specification in humans. Heterogeneity in differentiation potential amongst independently derived pluripotent stem cell lines is not unique to the germline, with multiple studies showing that hESC and hiPSC lines each have varying potentials for somatic cell differentiation [43–46]. In these studies, the underlying cause of this heterogeneity is thought to be due to differences in chromatin and DNA methylation in the self-renewing state [45, 46]. The underlying cause of variable germline competency also warrants further investigation.

A recent study using nine hiPSC lines cultured on StemFit also revealed that genes associated with primitive streak formation are associated with germline competency of hiPSCs in vitro [21]. Our study extends this observation by demonstrating that primitive streak gene expression is also associated with germline competency of hESCs, and that functionally, PGCLC competency can be attributed to the appropriate induction of *T* and *EOMES* downstream of TGFβ and WNT signaling. A recent study using hiPSCs revealed that T is not required for hPGCLC specification; therefore, even though *T* is induced and associated with PGCLC competency, *T* itself may not necessary for germline competency in vitro [17].

Unlike T, EOMES has no reported functional role in mouse, cyno, or porcine PGC specification. In mouse, EOMES is required for the cell migration in the primitive streak and *EOMES* loss-of-function mutants fail to undergo germ layer formation [47]. In the current study, we determined that EOMES is critical for human PGCLC specification where EOMES functions cell autonomously. Our RNA-seq data show that EOMES is expressed in iMeLCs but not in PGCLCs or PGCs. Therefore, we hypothesize that EOMES functions before the specification of SOX17 positive PGCs and is associated with germ cell competency of the epiblast/primitive streak. Whether EOMES regulates the migration of iMeLCs and whether the migration of iMeLCs is required for germ cell fate specification require further analysis.

Finally, our work touches on the potential origins of the human germline. In mouse embryos, PRDM1 positive PGC precursors are specified by BMP4 signaling to the Wnt3 primed posterior epiblast around the time of primitive streak formation [6, 15, 48]. In cyno embryos, PGCs were first identified prior to primitive streak formation in an embryonic cell layer called the amnion [3]. In porcine embryos, SOX17 positive PGCs were identified in prestreak epiblasts similar to the mouse [5]. In the current study, we discovered that signaling pathways and transcription factors associated with primitive streak formation (EOMES) regulate competency for SOX17 positive human PGCLC formation. This does not directly refute the origin of PGCLCs as being from the amnion, as the molecular identity of amnion cells is almost completely unknown. Instead, our work demonstrates that embryonic cells with the appropriate competency to respond to TGFβ and Wnt3 in order to induce EOMES are required for human PGCLC formation. Future studies using embryo attachment culture will be necessary to determine the origin of the

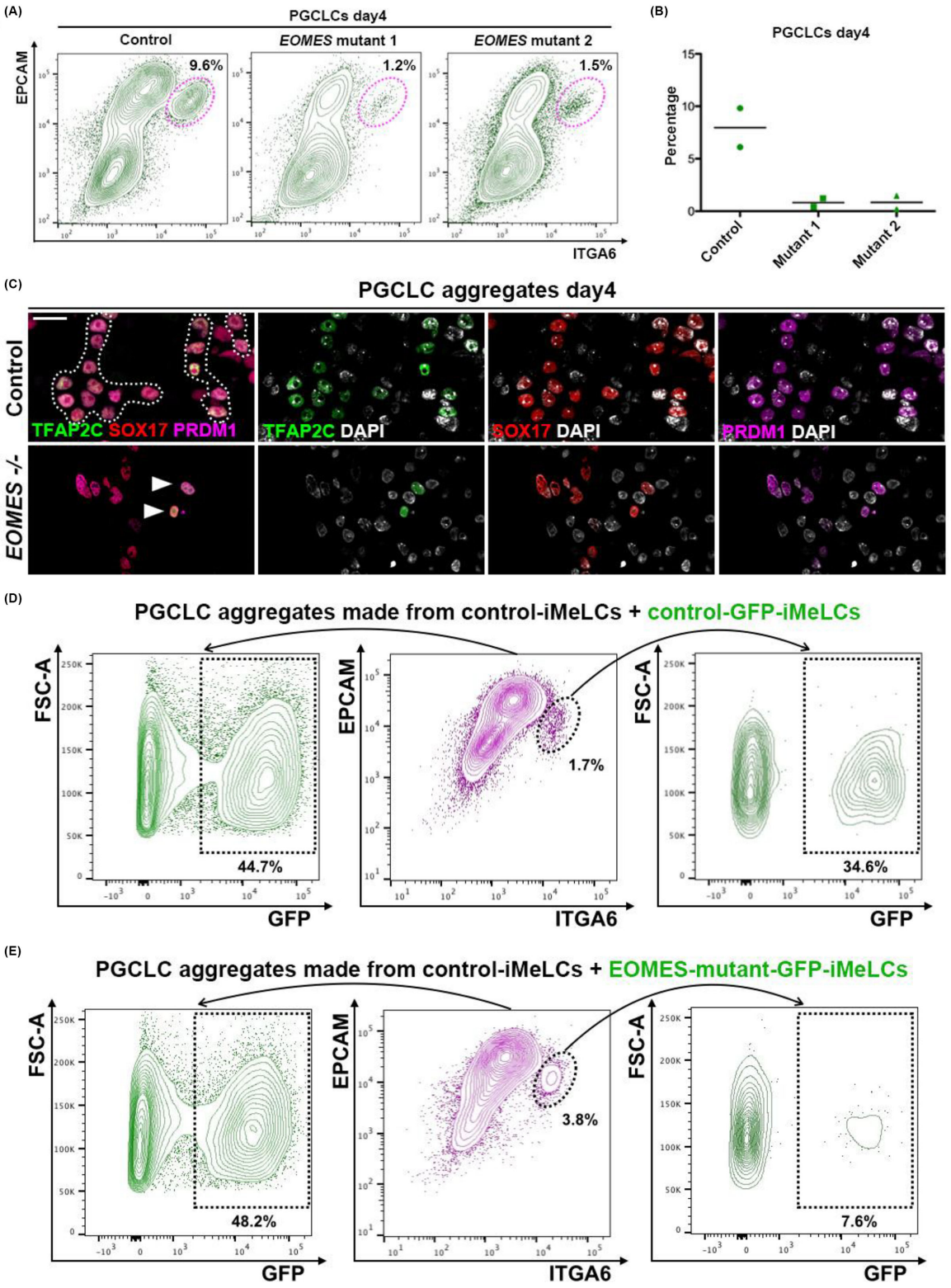

Figure 4. EOMES is required for PGCLC induction from hESCs. (A) Flow cytometry showing reduced PGCLC competency in *EOMES* mutant hESC line. PGCLCs were identified as ITGA6/EPCAM double positive cells. (B) Summary of PGCLC induction percentage from control and two different *EOMES* mutant hESC clones. (C) Control and *EOMES* mutant day 4 aggregates stained with TFAP2C (green), SOX17 (red), PRDM1 (purple), and DAPI (white). White dot outlines TFAP2C/SOX17/PRDM1 triple positive PGCLCs in control. White arrowheads point to rare TFAP2C/SOX17/PRDM1 triple positive PGCLCs in *EOMES* mutant. (D) Flow cytometry analysis of PGCLCs made from mixed iMeLCs (1:1 ratio) made from GFP negative and GFP positive wild-type hESCs. Left panel shows GFP positive cells in all live cells from PGCLC aggregates. Middle panel shows PGCLCs positive for ITGA6 and EPCAM. Right panel shows GFP positive PGCLCs. (E) Same analysis as (D) for PGCLCs made from mixed iMeLCs (1:1 ratio) made from GFP negative wild-type hESCs and GFP positive *EOMES* mutant hESCs.

human germline, but only if they emerge before the primitive streak given the ethical barrier that our field currently cannot cross.

Supplementary data

Supplementary data are available at *BIOLRE* online.

Figure S1. Germ cell marker expression and optimization of PGCLC induction from hESCs. (A) Human embryonic testis at day 115 section stained with ITGA6 (green), OCT4 (red), and DAPI (blue) by immunofluorescence. Yellow arrowhead points to an OCT4-positive germ cell that is also positive for ITGA6. Purple arrowhead points to an ITGA6-positive germ cell that is negative for OCT4. Scale bar: 10 μm. (B) Human embryonic testis at day 72 section stained with EPCAM (green), VASA (red), and DAPI (blue) by immunofluorescence. Yellow arrowhead points to a VASA-positive germ cell that is also positive for EPCAM. Purple arrowhead points to an EPCAM-positive germ cell that is negative for VASA. Scale bar: 10 μm. (C) Heat map showing the expression of germ cell, somatic cell, and pluripotency genes in transcriptomes shown in Figure 1E.

Figure S2. Generating ITGA6/EPCAM//TNAP/cKIT positive PGCLCs. (A and B) Flow cytometry of day 4 aggregates made with SCF (A) or without SCF (B) from UCLA2 (p13) and stained for ITGA6/EPCAM/TNAP/cKIT. (C) Real-time PCR of the sorted ITGA6/EPCAM positive PGCLCs made from UCLA2 (p13, p14) with SCF (green open columns) or without SCF (purple open columns) and compared to gene expression in ITGA6/EPCAM PGCs at day 72 postfertilization (Figure 1A and B) (green solid columns). Fold change is calculated relative to expression levels of each gene in the UCLA1 hESC line, which was given a value of 1.0. (D) Immunofluorescence of day 4 PGCLC aggregates from UCLA2 (p14) to examine colocalization of OCT4 (red), with cKIT, PRDM1, and TFAP2C (all green). Scale bar: 10 μm. (E) UHC of primed undifferentiated hESCs (UCLA1 p14, p15 and UCLA2 p13, p14), day 4 PGCLCs sorted by FACS using ITGA6/EPCAM (made from UCLA1 p14, p15 and UCLA2 p13, p14) with (plus) and without (minus) SCF, undifferentiated 4i cultured hESCs sorted with TNAP (WIS2) and day 4 PGCLCs sorted by FACS using TNAP/NANOS3-mCherry (made from WIS2). UHC was based on the expression of DEGs between hPGCs and H9 hESCs defined by Irie et al. [12] and Sasaki et al. [19]. U1 and U2 indicate UCLA1 and UCLA2, respectively. Gonadal germ cell libraries analyzed here are the same in Figure 1E. MS: minus SCF. PS: plus SCF. (F) PCA of transcriptomes in (E).

Figure S3. PGCLC induction from 18 pluripotent hESC lines derived at UCLA. (A) Morphology of human embryos used for derivation of hESC lines UCLA7, UCLA11, UCLA12, UCLA13, and UCLA15. All other hESC lines are reported elsewhere. (B) Karyotypes of hESC lines UCLA7, UCLA11, UCLA12, UCLA13, and UCLA15 (from left to right). All other hESC lines are reported elsewhere. (C) Representative images showing teratomas formed by injection of hESC lines UCLA7, UCLA11, UCLA12, UCLA13, and UCLA15 (from left to right) into the testes of SCID-beige mice. All other hESC lines are reported elsewhere. (D) FACS plots of day 4 PGCLCs (sorted with ITGA6/EPCAM) induced from 18 hESC lines through 24 h of iMeLC differentiation.

Figure S4. Evaluation of *T* and *EOMES* in different combination of cytokines and inhibitors and molecular information of *EOMES* mutant alleles. (A) *T* expression in the iMeLCs with different combinations of cytokines and signaling inhibitors. (B) *EOMES* expression in the iMeLCs with different combinations of cytokines and signaling inhibitors. (C) Molecular information of gRNAs for targeting *EOMES* and the resulting *EOMES* mutant alleles in the subline used in this study.

Table S1. List of antibodies used in this study.

Acknowledgments

The authors would like to thank Felicia Codrea and Jessica Scholes for FACS, Jinghua Tang for banking and culturing of the UCLA hESC lines, and Steven Peckman from the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research for critical assistance with human subject and embryonic stem cell review. Human conceptus tissue requests can be made to bdrl@u.washington.edu. SEJ is an investigator of the Howard Hughes Medical Institute.

Conflict of Interest: The authors have declared that no conflict of interest exists.

References

1. Magnusdottir E, Surani MA. How to make a primordial germ cell. *Development* 2014; **141**:245–252.
2. Tang WWC, Kobayashi T, Irie N, Dietmann S, Surani MA. Specification and epigenetic programming of the human germ line. *Nat Rev Genet* 2016; **17**:585–600.
3. Sasaki K, Nakamura T, Okamoto I, Yabuta Y, Iwatani C, Tsuchiya H, Seita Y, Nakamura S, Shiraki N, Takakuwa T, Yamamoto T, Saitou M. The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Dev Cell* 2016; **39**:169–185.
4. Clark AT, Gkountela S, Chen D, Liu W, Sosa E, Sukhwani M, Hennebold JD, Orwig KE. Primate primordial germ cells acquire transplantation potential by carnegie stage 23. *Stem Cell Reports* 2017; **9**: 329–341.
5. Kobayashi T, Zhang H, Tang WWC, Irie N, Withey S, Klisch D, Sybirna A, Dietmann S, Contreras DA, Webb R, Allegrucci C, Alberio R et al. Principles of early human development and germ cell program from conserved model systems. *Nature* 2017; **546**:416–420.
6. Ohinata Y, Payer B, O'carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, Saitou M, Surani MA. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 2005; **436**:207–213.
7. Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y, Saitou M. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* 2008; **40**:1016–1022.
8. Weber S, Eckert D, Nettersheim D, Gillis AJ, Schäfer S, Kuckenberg P, Ehlermann J, Werling U, Biermann K, Looijenga LH, Schorle H. Critical function of AP-2gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biol Reprod* 2010; **82**:214–223.
9. Nakaki F, Hayashi K, Ohta H, Kurimoto K, Yabuta Y, Saitou M. Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* 2013; **501**:222–226.
10. Magnúsdóttir E, Dietmann S, Murakami K, Günesdogan U, Tang F, Bao S, Diamanti E, Lao K, Gottgens B, Azim Surani M. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat Cell Biol* 2013; **15**:905–915.
11. Sugawa F, Arauzo-Bravo MJ, Yoon J, Kim K-P, Aramaki S, Wu G, Stehling M, Psathaki OE, Hubner K, Scholer HR. Human primordial germ cell commitment in vitro associates with a unique PRDM14 expression profile. *EMBO J* 2015; **34**:1009–1024.
12. Irie N, Weinberger L, Tang WWC, Kobayashi T, Viukov S, Manor YS, Dietmann S, Hanna JH, Surani MA. SOX17 is a critical specifier of human primordial germ cell fate. *Cell* 2015; **160**:253–268.
13. Lawson KA, Hage WJ. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* 1994; **182**:68–84; discussion 84–91.
14. Tam PP, Zhou SX. The allocation of epiblast cells to ectodermal and germline lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol* 1996; **178**:124–132.

15. Aramaki S, Hayashi K, Kurimoto K, Ohta H, Yabuta Y, Iwanari H, Mochizuki Y, Hamakubo T, Kato Y, Shirahige K, Saitou M. A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev Cell* 2013; **27**:516–529.
16. Ciruna BG, Rossant J. Expression of the T-box gene Eomesodermin during early mouse development. *Mech Dev* 1999; **81**:199–203.
17. Kojima Y, Sasaki K, Yokobayashi S, Sakai Y, Nakamura T, Yabuta Y, Nakaki F, Nagaoka S, Woltjen K, Hotta A, Yamamoto T, Saitou M. Evolutionarily distinctive transcriptional and signaling programs drive human germ cell lineage specification from pluripotent stem cells. *Cell Stem Cell* 2017; **21**:517.e5–532.e5.
18. Lawson KA, Dunn NR, Roelen BAJ, Zeinstra LM, Davis AM, Wright CVE, Korving JPWFM, Hogan BLM. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* 1999; **13**:424–436.
19. Sasaki K, Yokobayashi S, Nakamura T, Okamoto I, Yabuta Y, Kurimoto K, Ohta H, Moritoki Y, Iwatani C, Tsuchiya H, Nakamura S, Sekiguchi K et al. Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* 2015; **17**:178–194.
20. Kee K, Angeles VT, Flores M, Nguyen HN, Reijo Pera RA. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature* 2009; **462**:222–225.
21. Yokobayashi S, Okita K, Nakagawa M, Nakamura T, Yabuta Y, Yamamoto T, Saitou M. Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate. *Biol Reprod* 2017; **96**:1154–1166.
22. Diaz Perez SV, Kim R, Li Z, Marquez VE, Patel S, Plath K, Clark AT. Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. *Hum Mol Genet* 2012; **21**:751–764.
23. Patel S, Bonora G, Sahakyan A, Kim R, Chronis C, Langerman J, Fitz-Gibbon S, Rubbi L, Skelton RJP, Ardehali R, Pellegrini M, Lowry WE et al. Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep* 2017; **18**:54–67.
24. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; **339**:819–823.
25. Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M, Clark AT. 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* 2013; **15**:113–122.
26. Li Z, Yu J, Hosohama L, Nee K, Gkountela S, Chaudhari S, Cass AA, Xiao X, Clark AT. The Sm protein methyltransferase PRMT5 is not required for primordial germ cell specification in mice. *EMBO J* 2015; **34**:748–758.
27. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**:1105–1111.
28. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31**:166–169.
29. Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. A unique gene regulatory network resets the human germline epigenome for development. *Cell* 2015; **161**:1453–1467.
30. Tamura RN. Epithelial integrin alpha 6 beta 4: complete primary structure of alpha 6 and variant forms of beta 4. *J Cell Biol* 1990; **111**:1593–1604.
31. Trzpis M, McLaughlin PMJ, de Leij LMFH, Harmsen MC. Epithelial cell adhesion molecule. *Am J Pathol* 2007; **171**:386–395.
32. Gkountela S, Zhang KX, Shafiq TA, Liao W, Hargan-Calvopiña J, Chen P, Clark AT. DNA demethylation dynamics in the human prenatal germline. *Cell* 2015; **161**:1425–1436.
33. Loveland KL, Schlatt S. Stem cell factor and c-kit in the mammalian testis: lessons originating from Mother Nature's gene knockouts. *J Endocrinol* 1997; **153**:337–344.
34. Pesce M, Di Carlo A, De Felici M. The c-kit receptor is involved in the adhesion of mouse primordial germ cells to somatic cells in culture. *Mech Dev* 1997; **68**:37–44.
35. Richardson BE, Lehmann R. Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat Rev Mol Cell Biol* 2010; **11**:37–49.
36. Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, Yong J, Hu Y, Wang X, Wei Y, Wang W, Li R et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* 2015; **161**:1437–1452.
37. Cruse G, Beaven MA, Music SC, Bradding P, Gilfillan AM, Metcalfe DD. The CD20 homologue MS4A4 directs trafficking of KIT toward clathrin-independent endocytosis pathways and thus regulates receptor signaling and recycling. *Mol Biol Cell* 2015; **26**:1711–1727.
38. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 2005; **33**:W741–W748.
39. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 2013; **41**:W77–W83.
40. Funa NS, Schachter KA, Lerdrup M, Ekberg J, Hess K, Dietrich N, Honoré C, Hansen K, Semb H. β-Catenin regulates primitive streak induction through collaborative interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell* 2015; **16**:639–652.
41. Inman GJ. SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol Pharmacol* 2002; **62**:65–74.
42. Moon RT, Kohn AD, De Ferrari GV, Kaykas A. WNT and beta-catenin signalling: diseases and therapies. *Nat Rev Genet* 2004; **5**:691–701.
43. Feng Q, Lu S, Klimanskaya I, Gomes I, Kim D, Chung Y, Honig GR, Kim K, Lanza R. Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells* 2010; **28**:704–712.
44. Hu B-Y, Weick JP, Yu J, Ma L-X, Zhang X-Q, Thomson JA, Zhang S-C. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci USA* 2010; **107**:4335–4340.
45. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggan K et al. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 2011; **144**:439–452.
46. Butcher LM, Ito M, Brimpari M, Morris TJ, Soares FAC, Ährlund-Richter L, Carey N, Vallier L, Ferguson-Smith AC, Beck S. Non-CG DNA methylation is a biomarker for assessing endodermal differentiation capacity in pluripotent stem cells. *Nat Commun* 2016; **7**:10458.
47. Russ AP, Wattler S, Colledge WH, Aparicio SA, Carlton MB, Pearce JJ, Barton SC, Surani MA, Ryan K, Nehls MC, Wilson V, Evans MJ. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* 2000; **404**:95–99.
48. Kurimoto K, Saitou M. Mechanism and reconstitution in vitro of germ cell development in mammals. *Cold Spring Harb Symp Quant Biol* 2015; **80**:147–154.