Stem Cell Reports
Article

ISSCR

OPEN ACCESS

PDGFRα$^+$ Cells in Embryonic Stem Cell Cultures Represent the In Vitro Equivalent of the Pre-implantation Primitive Endoderm Precursors

Antonio Lo Nigro,[1,7,8,*] Anchel de Jaime-Soguero,[1,8] Rita Khoueiry,[1] Dong Seong Cho,[2] Giorgia Maria Ferlazzo,[1] Ilaria Perini,[1] Vanesa Abon Escalona,[3,4] Xabier Lopez Aranguren,[5] Susana M. Chuva de Sousa Lopes,[6] Kian Peng Koh,[1] Pier Giulio Conaldi,[7] Wei-Shou Hu,[2] An Zwijsen,[3,4] Frederic Lluis,[1,9] and Catherine M. Verfaillie[1,9]

[1]Department of Development and Regeneration, Stem Cell Biology and Embryology, KU Leuven Stem Cell Institute, Herestraat 49, Onderwijs en Navorsing 4, Box 804, 3000 Leuven, Belgium
[2]Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue Southeast, Minneapolis, MN 55455, USA
[3]VIB Center for the Biology of Disease, 3000 Leuven, Belgium
[4]KU Leuven Department of Human Genetics, 3000 Leuven, Belgium
[5]Cell Therapy Program, Foundation for Applied Medical Research, University of Navarra, 31008 Pamplona, Spain
[6]Department of Anatomy and Embryology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, the Netherlands
[7]Ri.Med Foundation, Department of Laboratory Medicine and Advanced Biotechnologies, IRCCS-ISMETT (Istituto Mediterraneo per i Trapianti e Terapie ad Alta Specializzazione), Via Tricomi 5, 90127 Palermo, Italy
[8]Co-first author
[9]Co-senior author
*Correspondence: antoniolonigro83@gmail.com
http://dx.doi.org/10.1016/j.stemcr.2016.12.010

SUMMARY

In early mouse pre-implantation development, primitive endoderm (PrE) precursors are platelet-derived growth factor receptor alpha (PDGFRα) positive. Here, we demonstrated that cultured mouse embryonic stem cells (mESCs) express PDGFRα heterogeneously, fluctuating between a PDGFRα+ (PrE-primed) and a platelet endothelial cell adhesion molecule 1 (PECAM1)-positive state (epiblast-primed). The two surface markers can be co-detected on a third subpopulation, expressing epiblast and PrE determinants (double-positive). In vitro, these subpopulations differ in their self-renewal and differentiation capability, transcriptional and epigenetic states. In vivo, double-positive cells contributed to epiblast and PrE, while PrE-primed cells exclusively contributed to PrE derivatives. The transcriptome of PDGFRα$^+$ subpopulations differs from previously described subpopulations and shows similarities with early/mid blastocyst cells. The heterogeneity did not depend on PDGFRα but on leukemia inhibitory factor and fibroblast growth factor signaling and DNA methylation. Thus, PDGFRα$^+$ cells represent the in vitro counterpart of in vivo PrE precursors, and their selection from cultured mESCs yields pure PrE precursors.

INTRODUCTION

Totipotency is the capacity to form an entire organism, including embryonic and extraembryonic tissues. In mouse, totipotency lasts from fertilization at embryonic day (E)0 until the morula stage (~E2.5). Loss of totipotency, early in pre-implantation development, is accompanied by segregation of the first lineage: the outer trophectoderm (TE) that separates from the inner cell mass (ICM). At implantation (~E4.5), the ICM further generates two distinct layers: the epiblast and the primitive endoderm (PrE, also known as hypoblast) (Arnold and Robertson, 2009). At this stage, lineage identities are dictated by the expression of specific transcription factors (TFs). The pluripotent epiblast fate is induced by the expression of *Oct4*, *Nanog*, and *Sox2* (Wicklow et al., 2014; Yamanaka et al., 2010); the segregated PrE layer is positive for *Oct4*, *Gata4*, *Gata6*, *Sox7*, and *Sox17*, whereas the cells of the TE express *Cdx2* (Artus et al., 2011; Plusa et al., 2008). At earlier stages, these determinants are not specific: in the morula, embryonic and extraembryonic TFs are co-expressed in all blastomeres (Bessonnard et al., 2014; Dietrich and Hiiragi, 2007; Guo et al., 2010; Ohnishi et al., 2014; Schrode et al., 2014).

Proceeding with development, the epiblast forms all embryonic tissues but also the extraembryonic mesoderm of the visceral yolk sac, the chorion, the allantois, and the amnion. The PrE subsequently gives rise to the parietal endoderm (PE) of the transient parietal yolk sac and the visceral endoderm (VE). The VE consists of embryonic and extraembryonic VE. The extraembryonic VE, together with extraembryonic mesoderm, forms the visceral yolk sac, while the embryonic VE is necessary for correct anterior-posterior patterning of the embryo. In addition, recent findings suggest that embryonic VE also contributes to the gut (Kwon et al., 2008). The TE forms trophoblast giant cells, the extraembryonic ectoderm and its derivatives, the ectoplacental cone, and the chorionic ectoderm. TE is necessary for implantation of the conceptus and exchange of products between the maternal and fetal circulation.

Mouse embryonic stem cell (ESC) lines are derived from the ICM of developing blastocysts at ~E3.5 (Evans and Kaufman, 1981; Martin, 1981). ESC lines capture many

features of the epiblast and are defined as pluripotent because they can differentiate into the three definitive germ layers of the embryo when injected in recipient blastocysts or aggregated with morulas. In addition, pluripotent ESC lines can also generate trophoblast (Hayashi et al., 2010) and PrE cell types in vitro (i.e., extraembryonic endodermal cells [XENs]) (Kunath et al., 2005; Niakan et al., 2013), aside from cells of the three germ layers of the embryo. There is also evidence that ESCs rarely contribute to extraembryonic lineages in vivo (Beddington and Robertson, 1989). Taken together, these data indicate that ESC cultures contain precursors of extraembryonic lineages.

Traditionally, ESCs were derived and cultured in the presence of leukemia inhibitory factor (LIF) and either bone morphogenetic protein 4 (BMP4) or fetal bovine serum (BMP4/L or FBS/L) (Ying et al., 2003a). Under such conditions, ESC cultures are heterogeneous and contain metastable and fluctuating subpopulations, resembling later (post-implantation epiblast) or earlier (two-cell stage) developmental stages (Hayashi et al., 2008; Macfarlan et al., 2012). Recently, efficient and clonal derivation from ICM cells (Boroviak et al., 2014) was reported by using a defined medium containing two inhibitors of MEK and GSK3β kinases together with LIF (2i/L). ESC lines cultured in 2i/L maintain a less heterogeneous "naive" ground state (Marks et al., 2012; Ying et al., 2008).

Early in development, PDGFRα has a relatively weak but well visible expression in all blastomeres until it becomes stronger in PrE-committed cells around E3.75 (around 64 cells) (Artus et al., 2011; Grabarek et al., 2012; Plusa et al., 2008). Here, we demonstrate that PDGFRα$^+$ cells can also be identified in undifferentiated ESC cultures. The PDGFRα$^+$ subpopulations show a unique PrE-primed molecular and epigenetic signature, which is reflected by functional in vitro and in vivo differences when compared with the epiblast counterpart (PECAM1$^+$). Despite these differences, the transcriptome of PDGFRα$^+$ cells displays similarities with naive ESCs and with early/mid blastocyst cells. These findings suggest that PDGFRα$^+$ cells are the equivalent of the in vivo PrE (hypoblast) precursors present at the pre-implantation stage.

RESULTS

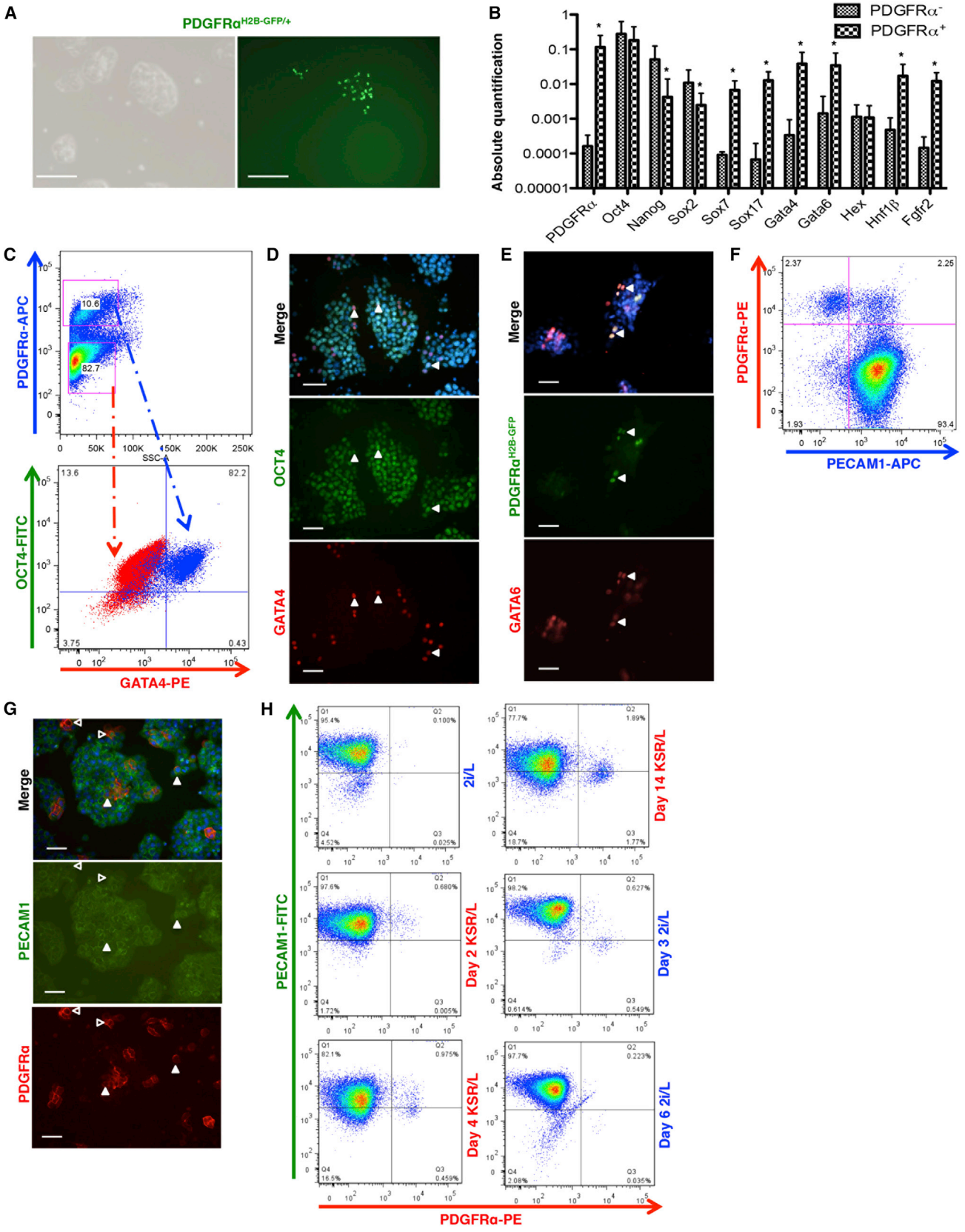
ESC Cultures Contain a PDGFRα$^+$ Subpopulation When Cultured without 2i

Expression of PDGFRα has been reported in differentiating ESCs and in XEN cells, but not in undifferentiated ESC lines. Here, we investigated its expression by using a *Pdgfrα*$^{H2B\text{-}GFP/+}$ reporter line (Hamilton et al., 2003) in which the H2B-GFP fusion protein tracks its presence. GFP$^+$ cells were detected within colonies of ESC lines, cultured in LIF and knockout serum replacement (KSR/L) (Bryja et al., 2006) (Figure 1A). The comparison between GFP$^+$ and negative cells by qRT-PCR, upon separation by fluorescence-activated cell sorting (FACS), showed that *Oct4* transcript levels in PDGFRα$^+$ cells were similar to those detected in PDGFRα$^-$ cells, while *Nanog* and *Sox2* transcripts were expressed at lower levels (Figure 1B). Transcript levels of genes associated with early extraembryonic fate (*Gata4*, *Gata6*, *Sox7*, *Sox17*, *Hnf1β*, and *Fgfr2*) were higher in the PDGFRα$^+$ fraction (Figure 1B). Although cells with a PrE profile (Canham et al., 2010) have been described as *Hex*$^+$, the *Hex* transcript levels were identical in the two fractions (Figure 1B).

We next stained E14 and R1 ESCs cultured in KSR/L with antibodies against PDGFRα, OCT4, and GATA4. This confirmed the presence of a PrE-primed subpopulation (Figure 1C, top plot), as ±80% of the PDGFRα$^+$ cells co-expressed OCT4 and GATA4 (Figure 1C, bottom plot and 1D), differently from PDGFRα$^-$ cells, which expressed OCT4 only. We also found co-staining for PDGFRα and GATA6 (Figure 1E), using a *Sox17*$^{:GFP/+}$ ESC line between SOX17 and PDGFRα (>50% of PDGFRα$^+$ were GFP$^+$, Figure S1A), and between SOX17 and OCT4 (Figure S1B). The molecular identity (OCT4, GATA4, GATA6, and SOX17) of PDGFRα$^+$ cells strongly resembles the pre-implantation (∼E3.75) PrE precursor (Artus et al., 2011).

During the transition from morula to early blastocyst stage, cells co-express markers that later become specific for either epiblast or PrE. We tested whether PECAM1, a marker of epiblast in ICM and ESCs, was co-expressed with PDGFRα, to understand if expression of epiblast and PrE surface markers was mutually exclusive in vitro. We identified three different subpopulations: PECAM1$^+$/PDGFRα$^-$ (epiblast-primed), PECAM1$^+$/PDGFRα$^+$ (double-positive), and PECAM1$^-$/PDGFRα$^+$ (PrE-primed) cells (Figures 1F, 1G, and S1C). Consistently, in the *Sox17*$^{:GFP/+}$ ESC line, a subpopulation of PECAM1$^+$ cells was also GFP$^+$ (Figures S1D and S1E).

Previous reports suggested that culture in 2i/L maintains the expression of early endodermal genes (Canham et al., 2010; Marks et al., 2012). To test this, we cultured the *Pdgfrα*$^{H2B\text{-}GFP/+}$ ESCs for 3 days in 2i/L. As shown in Figure S2A, this resulted in a loss of GFP$^+$ cells, a decrease of extraembryonic transcripts (*Gata4, Gata6, Sox7, Sox17, FoxA2*, and *Hnf1β*), and an increase of *Nanog* (Figure S2B); whereas an increase of extraembryonic transcripts levels was seen upon 2i withdrawal (Figure S2C). To confirm the effect of naive culture conditions, we adapted the R1 ESC line to 2i/L for 3 weeks (Figure 1H). In 2i/L, the PDGFRα$^+$ subpopulations were strongly reduced. Subsequent withdrawal of 2i leads to the appearance of the double-positive subpopulation in 2 days and of the PrE-primed



(legend on next page)

subpopulation in 4 days (Figure 1H, left plots). When culture conditions were again switched to 2i/L, the PDGFRα$^+$ subpopulations almost completely disappeared in 6 days (Figure 1H, right plots). To determine whether the loss of PDGFRα$^+$ cells was due to decreased proliferation or increased apoptosis of the PDGFRα$^+$ cells, we performed triple intracellular staining for PECAM1, PDGFRα, and either KI67 (proliferation marker) or active CASPASE3 (apoptotic marker). This analysis showed that, under 2i/L, the proliferation of PDGFRα$^+$ cells decreased (Figure S2D) without a significant increase in cell death (Figure S2E), demonstrating that the faster proliferating epiblast-primed subpopulation became predominant and took over the culture.

Molecular and Functional Differences of the PDGFRα$^+$ Subpopulations

As we could co-detect epiblast and PrE surface markers (Figure 1F), we confirmed the expression of epiblast and PrE TFs at the single-cell level by performing triple intracellular staining for OCT4, GATA4, and NANOG (Figure 2A). This showed that ~8% of the cells co-expressing OCT4 and GATA4 (top left plot) also expressed NANOG (top right plot).

While comparing the different subpopulations, we detected in epiblast-primed cells high levels of the pluripotency transcripts (*Oct4*, *Nanog*, *Sox2*, and *Esrrβ*) as well as proteins (OCT4, NANOG, and SOX2), but low/no expression of PrE-related genes (Figures 2B and 2C). By contrast, in the PrE-primed cells, *Oct4* transcripts and protein could be detected, whereas NANOG and SOX2 could not. Moreover, expression of markers specific for an extraembryonic (Figure 2B) but not of post-implantation epiblast fate (*Fgf5*, *T*, *Nodal*, *Nr0b1*, and *Otx2*; Figure S3A) (Brons et al., 2007) were significantly higher in the PrE-primed cells than in the other two-cell populations. Double-positive cells had an intermediate phenotype with respect to both single positive subpopulations.

To further characterize the three cell populations, they were isolated by FACS and subjected to in vitro functional tests. First, we cultured them at clonal density in KSR/L medium. In contrast to epiblast-primed and double-positive cells, PrE-primed cells poorly re-adhered to gelatin-coated plastic and rarely formed ESC colonies; they grew as single cells, resembling XEN cells (Kunath et al., 2005) and stained positive for alkaline phosphatase (Figure 2D). Time course FACS analysis showed that epiblast-primed and double-positive cells re-established the initial heterogeneity in a week when cultured in KSR/L, differently from PrE-primed cells, which strongly maintained a bias for the seeded subpopulation (Figures 2E and S3B), even when replated in 2i medium (Figure S3C).

We also tested whether PrE-primed sorted cells could be propagated in a stable and pure form by culturing them in medium that allows the derivation of OCT4$^+$/GATA4$^+$ PrE lines from rat blastocysts (Lo Nigro et al., 2012). However, prolonged culture (>3weeks) of PrE-primed sorted cells resulted in a mixture of cells with epiblast and PrE morphology and TFs (data not shown).

Second, we compared their differentiation potential into definitive endodermal by culturing them with Wnt3a and Activin A (Sancho-Bru et al., 2011). Time course analysis showed that *Goosecoid, Eomes*, and *Mixl1* could be detected in epiblast-primed and double-positive cells, but not in PrE-primed cells, while *T* was upregulated specifically in epiblast-primed progeny (Figures 3A and 3B). Differently, *Sox17*, *Sox7*, *Foxa2*, and *Cxcr4* (markers for definitive endoderm and for PrE-derivatives) were expressed from the beginning of the differentiation in PrE-primed and double-positive cells but not in epiblast-primed cells, wherein these markers were only upregulated at later stages. We also found that PDGFRα$^+$ subpopulations fail to generate

Figure 1. Undifferentiated ESC Cultures Contain PDGFRα-Expressing Cells

(A) Bright field picture and GFP expression in *Pdgfrα*$^{H2B\text{-}GFP/+}$ ESC lines. Scale bar, 100 μm.

(B) qRT-PCR analysis for embryonic and extraembryonic markers in *Pdgfrα*$^{H2B\text{-}GFP\,+/-}$ subpopulations. Data are presented as means ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *p < 0.05, t test.

(C) FACS analysis on E14 ESC lines for the expression of PDGFRα (top plot) and OCT4/GATA4 (bottom plot). The red cloud represents PDGFRα$^-$ cells, while the blue cloud represents PDGFRα$^+$ cells, n = 3. The gating strategy was based on isotype controls.

(D) Immunostaining analysis for OCT4 and GATA4 on R1 ESC line. Arrowheads indicate cells co-expressing OCT4 and GATA4. Scale bar, 50 μm, n = 3.

(E) Immunostaining analysis for GATA6 on *Pdgfrα*$^{H2B\text{-}GFP/+}$ ESC line. Arrowheads indicate cells co-expressing PDGFRα$^{H2B-GFP}$ and GATA6. Scale bar, 50 μm, n = 3.

(F) Representative FACS analysis for PDGFRα and PECAM1 on the R1 line, n = 3. For isotype controls, gating strategies, and sorting purities, see Figure S1C.

(G) Immunostaining for PDGFRα and PECAM1 on the R1 line. Empty arrowheads indicate cells expressing only PDGFRα, full arrowheads indicate cells co-expressing PDGFRα and PECAM1. Scale bar, 50 μm, n = 3.

(H) Representative time course FACS analysis for PDGFRα and PECAM1 on the R1 line in 2i/L and KSR/L, n = 3. Gating strategy was based on isotype controls.

See also Figures S1 and S2.

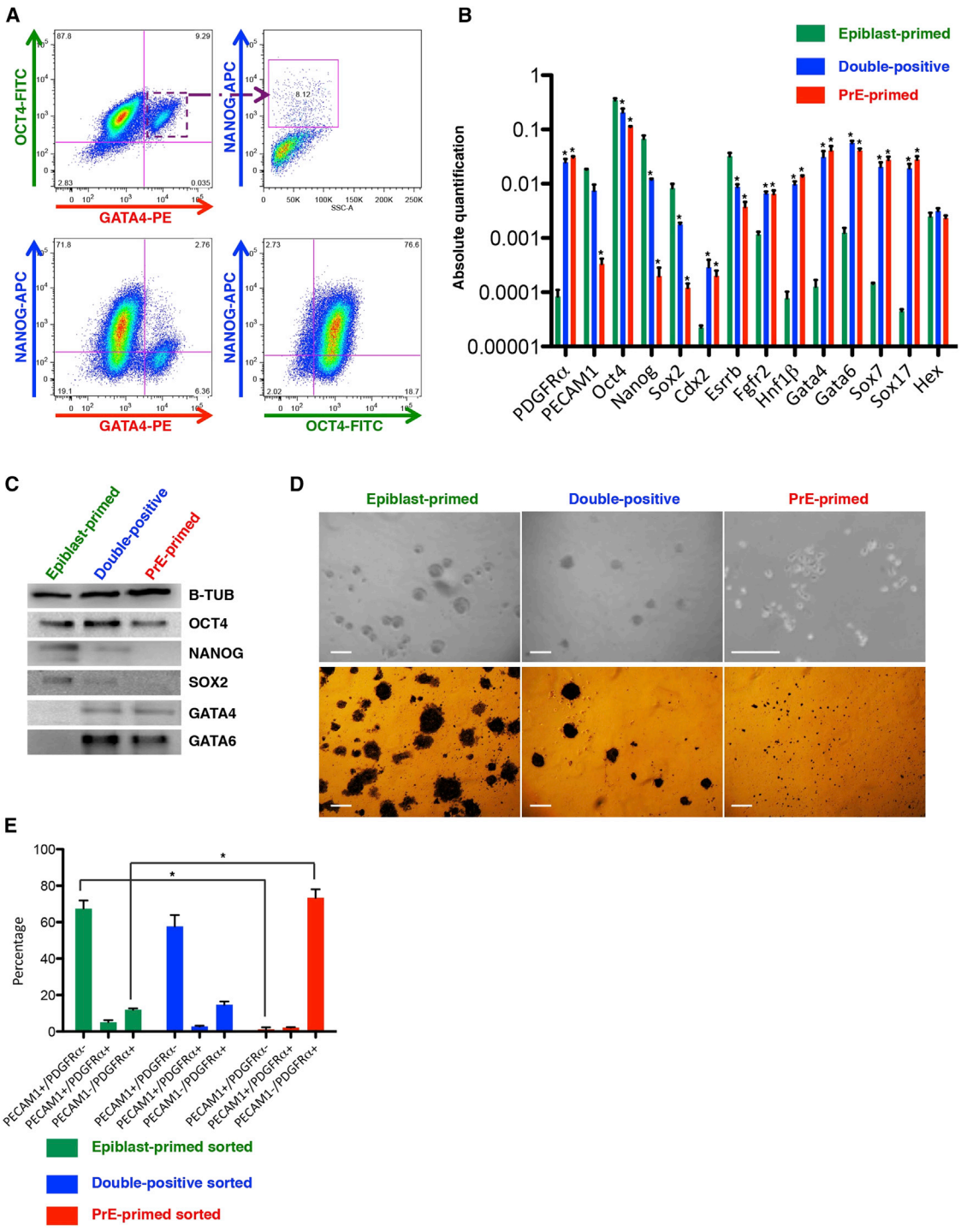


Figure 2. ESC Cultures Contain Different PDGFRα$^+$ Subpopulations

(A) Representative intracellular FACS analysis for OCT4, GATA4, and NANOG on the R1 line, n = 3. The gating strategy was based on isotype controls.

(B) qRT-PCR analysis for embryonic and extraembryonic markers in the three subpopulations. Data are presented as means ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *p < 0.05, t test.

(C) Representative western blot of three independent experiments for OCT4, NANOG, SOX2, GATA4, and GATA6 on sorted cells. B-TUBULIN was used as normalizer.

(legend continued on next page)

mesendoderm, suggesting a preferential differentiation toward PE/VE cell types. Similarly, upon induction of neuroectodermal lineage (Ying et al., 2003a), neural precursors were only detected in epiblast-primed progeny (arrows, Figure S3D). Double-positive and PrE-primed subpopulations formed vacuolated structures (empty arrows, Figure S3D), resembling differentiating XEN cells (Kunath et al., 2005).

Third, we assessed the capacity of the three cell populations to generate extraembryonic cell types in trophoblast stem cell (TSC) medium, also shown to support the derivation of XEN cells (Niakan et al., 2013). PrE-primed cells but not the other two-cell populations formed XEN-like colonies, positive for GATA6 and LAMININ-β2 and expressing PrE transcripts (Figures 3C and 3E). Although ESCs are not thought to be capable of generating TSCs without genetic manipulation, we evaluated the presence of putative trophoblast cell progeny, i.e., $CDX2^{+}GATA6^{-}$ cells (Figure 3D), as described (Morgani et al., 2013). Trophoblast-like progeny was only detected in epiblast-primed and double-positive cells cultures, as confirmed by qRT-PCR for *Cdx2*, *Gata3*, and *Krt7* (Figure 3E).

Different Epigenetic State of PDGFRα^{+} Subpopulations

DNA methylation is dispensable for the growth of extraembryonic but not of embryonic tissues (Sakaue et al., 2010); moreover extraembryonic tissues have a lower level of DNA methylation than their embryonic equivalent (Rossant et al., 1986). We therefore compared the expression of DNA methylation and hydroxymethylation genes in the three subpopulations. Levels of *Dnmt1*, *Dnmt3l*, and *Tet1* transcripts were significantly lower in PrE-primed cells compared with the other fractions (Figure 4A), suggesting a lower level of 5-mC and 5-hmC in PDGFRα^{+} cells. In addition, the promoter of intracisternal A-particle (IAP), which has repetitive elements with ±1,000 copies in the *Mus musculus* genome, was significantly less methylated in PDGFRα^{+} cells (±55%) compared with the epiblast-primed cells (±87%, Figure 4B). Accordingly, the genome-wide levels of 5-mC and 5-hmC in the genomic DNA were lower in PrE-primed cells compared with the other subpopulations (Figures 4C and 4D).

The in vitro model described here also reflects these crucial differences in methylation between embryonic and extraembryonic tissues. A similar hypomethylated state has been reported for naive ESCs, where 2i reduces DNA methylation by increasing *Prdm14* (Leitch et al., 2013). However, *Prmd14* and other genes involved in primordial germ cells specification/imprinting (*Dppa3*, *Dazl*, and *Prdm1*) were lower in PDGFRα^{+} subpopulations (Figure 4E), suggesting that DNA hypomethylation depends on other mechanisms.

Finally, we compared the transcript levels of genes involved in chromatin regulation. Polycomb repressive complex (PRC)-1 and -2 and their histone modifications are crucial for the dynamic equilibrium and the plasticity of ESCs by acting as transcriptional repressors (Boyer et al., 2006). Of note, RNA sequencing (RNA-seq) analysis showed remarkable differences between the three subpopulations for the expression of these epigenetic regulators (Figure 4E). Compared with the epiblast-primed fraction, the PDGFRα^{+} subpopulations expressed significantly lower levels of *Kdm2b*, *Jarid2* (which respectively recruit PRC1 and PRC2 complex to chromatin) and of *Ezh2*, *Eed*, and *Suz12* (PRC2 components), suggesting a lower level of H3K27 methylation, known to be reduced in extraembryonic cell types (Alder et al., 2010; Rugg-Gunn et al., 2010).

The Developmental Potential of PDGFRα^{+} Cells Reflects Their Different Molecular Identity

When ESCs are used to generate chimeras, chimerism is detected in the epiblast lineage that gives rise to all embryonic and to extraembryonic mesodermal tissues. However, ESC progeny has also been described to contribute very sporadically to TE or PrE-derived extraembryonic lineages (Beddington and Robertson, 1989). To compare their developmental potential, we injected GFP^{+} ESCs in recipient blastocysts after FACS sorting of the three subpopulations (Table S1). As expected, epiblast-primed cells efficiently colonized epiblast-derived tissues with high degrees of chimerism (Figure 5A) but not the TE/PrE-derived extraembryonic tissues. Injection of the double-positive subpopulation resulted in chimerism in the embryo proper (Figure 5B) as well as in the VE (Figure 5C) and PE (Figure 5D). PrE-primed cells contributed to both VE (Figure 5E) and PE (Figure 5F) but not to epiblast/TE derivatives. The behavior of the PrE-primed subpopulation differs from that of the *Hex*$^{+}$ cells (Canham et al., 2010), which showed a low contribution (10%) to PrE-derived tissues while still colonizing the embryo.

PDGFRα^{+} cells have a distinct molecular identity, which is further reflected by different developmental potential in vivo.

PDGFRα^{+} Subpopulations Have a Unique Expression Profile that Resembles Early/Mid Blastocyst Cells

To investigate genome-wide differences/similarities between the three subpopulations, we performed RNA-seq.

(D) Bright field pictures and alkaline phosphatase staining on sorted subpopulations. Scale bar, 100 μm.

(E) Percentage of each subpopulation 1 week after their respective sorting, calculated from three independent experiments, *$p < 0.05$, t test.

See also Figure S3.

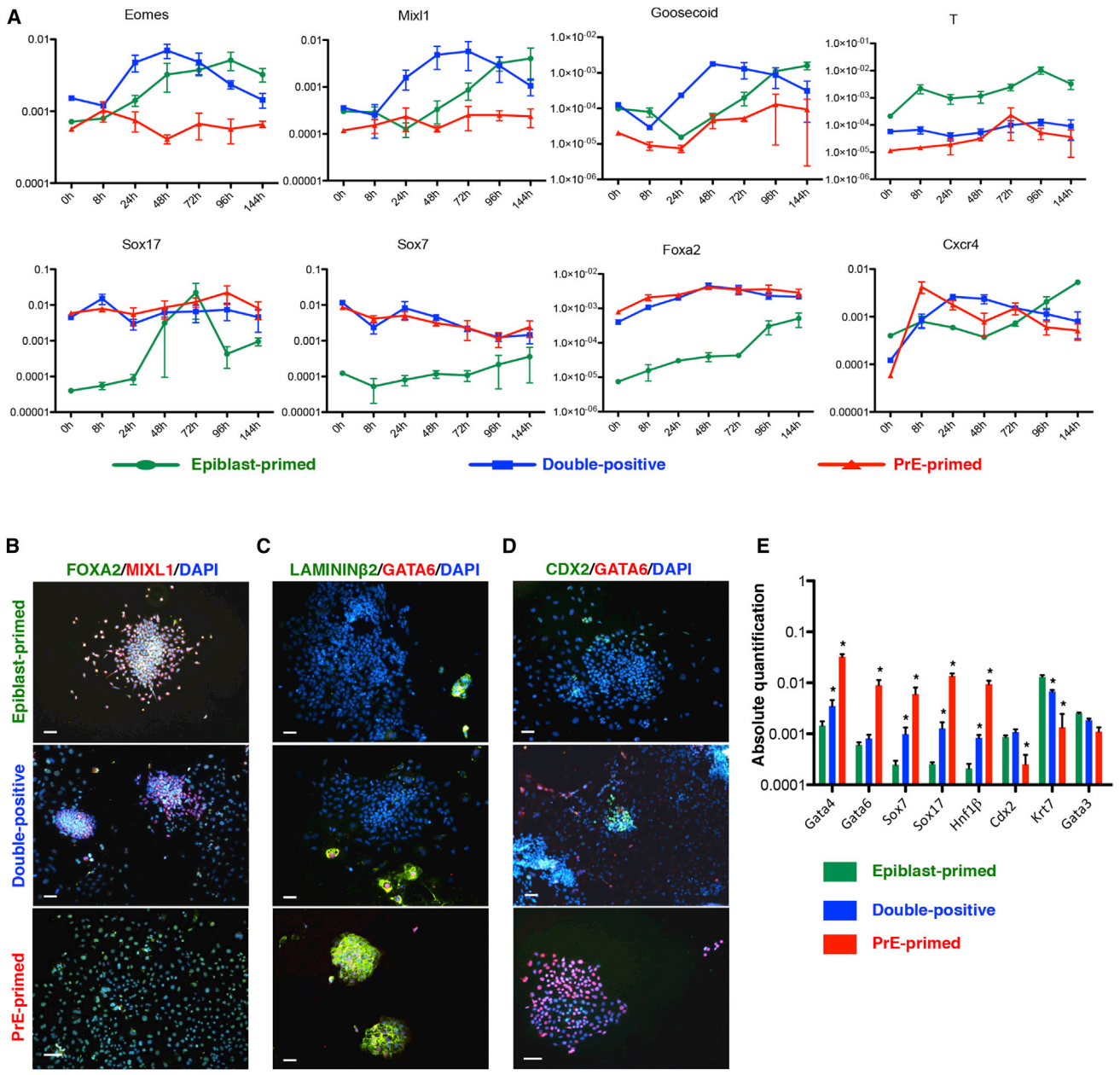


Figure 3. In Vitro Functional Differences of the Three Subpopulations

(A) qRT-PCR time course analysis for the indicated genes upon Wnt3a/Activin A treatment. Data are represented as means ± SEM of each transcript from three independent experiments (normalized to *β-Actin*).

(B) Immunostaining analysis for FOXA2 and MIXL1 on sorted subpopulations at day 4. Scale bar, 50 μm, n = 3.

(C) Immunostaining analysis for LAMININ-β2 and GATA6 on sorted subpopulations after 7 days in TSC medium. Scale bar, 50 μm, n = 3.

(D) Immunostaining analysis for CDX2 and GATA6 on the different sorted subpopulations after 7 days in TSC medium. Scale bar, 50 μm. n = 3.

(E) qRT-PCR analysis for XEN and trophoblast-related markers, upon sorting and differentiation. Data are presented as means ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *$p < 0.05$, t test.

See also Figure S3.

The comparison demonstrated that 292 genes were more than 2-fold differentially expressed in epiblast-primed cells, 41 in double-positive cells, and 2,131 in the PrE-primed subpopulation (Figure 6A and Table S2). The double-positive cells more closely resemble the PrE-primed state rather than the epiblast-primed state (Figures 6A

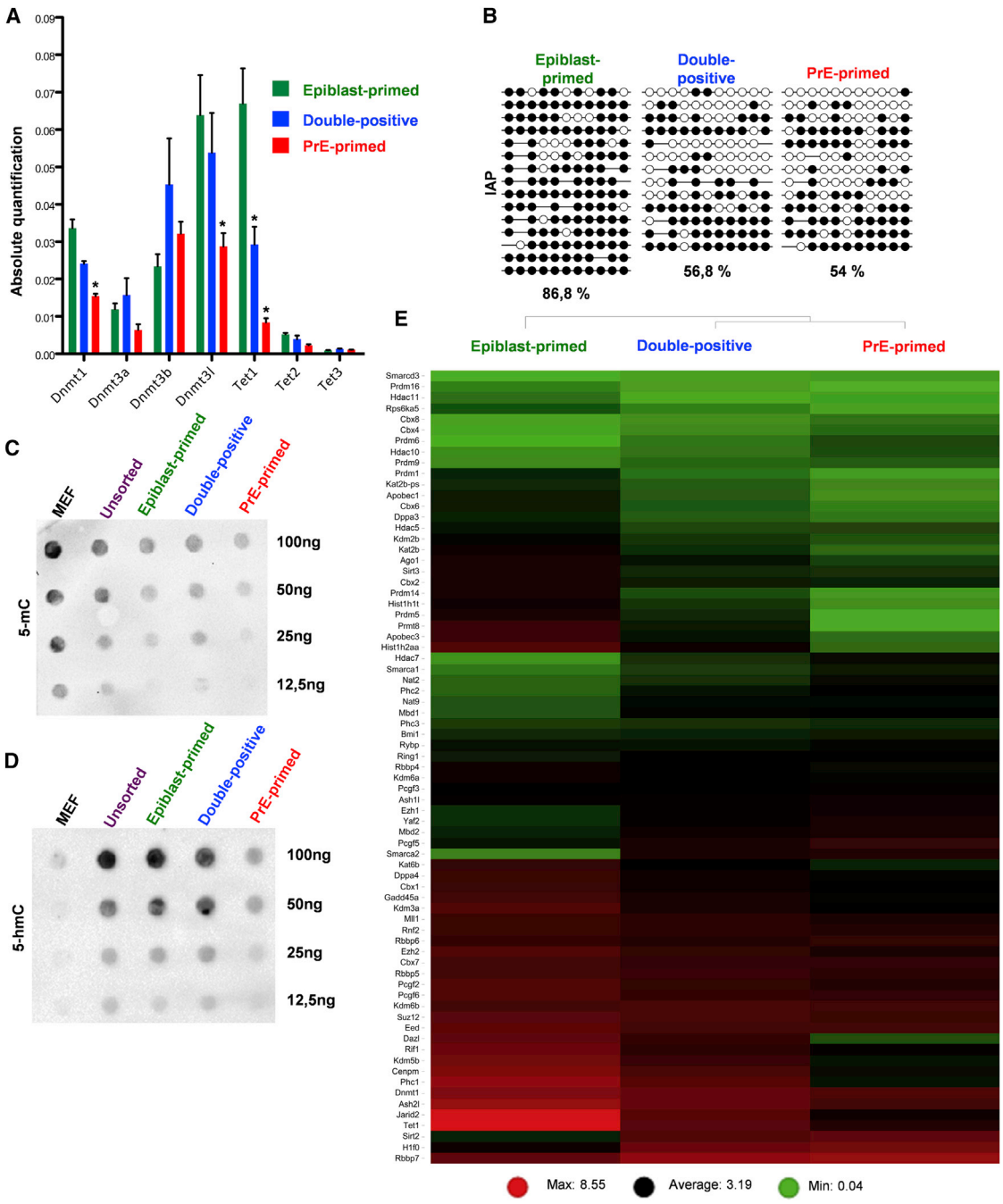


Figure 4. Different Epigenetic State of PDGFRα^+ Subpopulations

(A) qRT-PCR analysis for DNA methylation/hydroxymethylation genes upon sorting of the respective subpopulations. Data are presented as means ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *p < 0.05, t test.

(B) Bisulfite sequencing of IAP sequences. Open circles, unmethylated; closed circles, methylated.

(C) Representative dot blot for global 5-mC, from three independent experiments. Mouse embryonic fibroblasts (MEF) were used as control as they contain high levels of 5-mC and low levels of 5-hmC.

(D) Representative dot blot for global 5-hmC from three independent experiments. MEFs were used as control as they contain low levels of 5-hmC.

(E) Heatmap of epigenetic regulators on sorted subpopulations based on RNA-seq data. Transcript levels are based on FPKM (fragments per kilobase of exon per million fragments mapped). Red and green represent high and low gene expression, respectively.

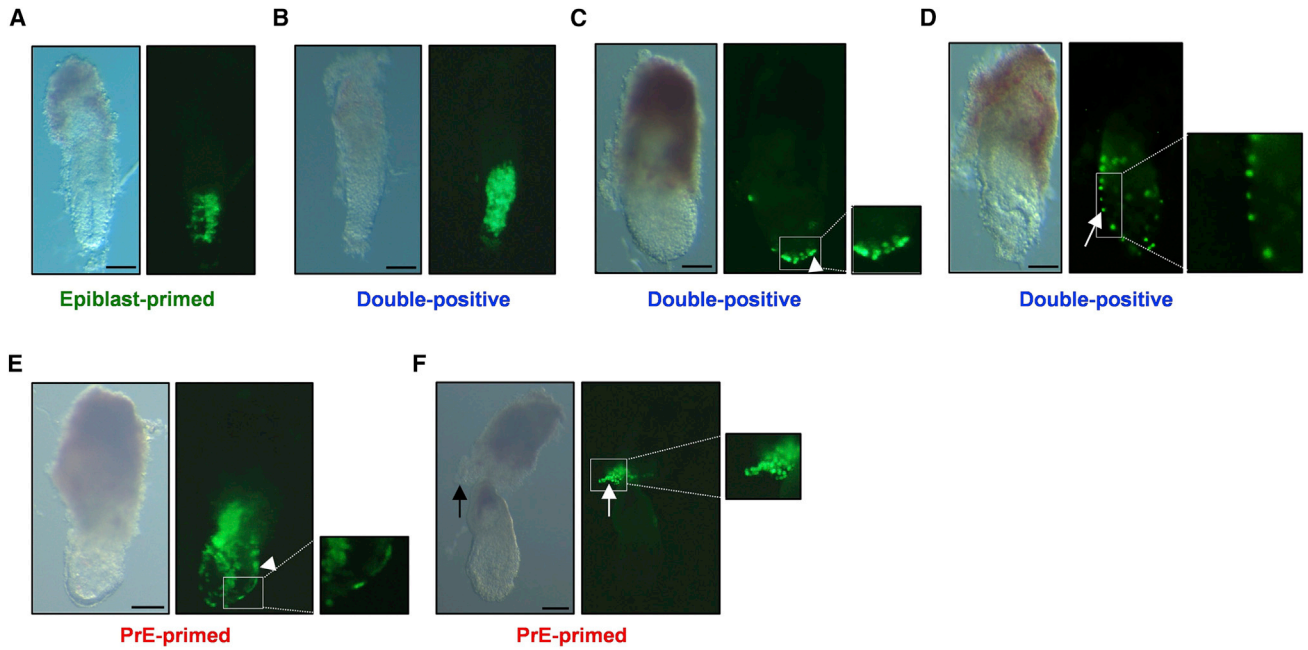


Figure 5. In Vivo Comparison of Developmental Potential

GFP$^+$ ESCs were FACS sorted for PECAM and/or PDGFRα and injected into recipients blastocysts. Scale bar, 100 μm.
(A) E6.5 chimeric embryo generated from epiblast-primed cells showing the contribution to the embryo proper.
(B) E6.5 chimeric embryo generated from double-positive cells showing the contribution to the embryo proper.
(C) E6.5 chimeric embryo generated from double-positive cells showing the contribution to the VE (arrowhead).
(D) E6.5 chimeric embryo generated from double-positive cells showing the contribution to the PE (arrow).
(E) E6.5 chimeric embryo generated from PrE-primed cells showing the contribution to the VE (arrowhead).
(F) E6.5 chimeric embryo generated from PrE-primed cells showing the contribution to the PE (arrow).
See also Table S1.

and S4B). As PDGFRα$^+$ subpopulations appear to have a PrE molecular phenotype, we compared the sorted subpopulations between each other but also with XEN isolated from embryo (eXEN) or converted from ESCs (cXEN), upon Activin A/retinoic acid treatment (Cho et al., 2012). Core and naive pluripotency genes (*Nanog*, *Sox2*, *Esrrb*, *Klf2*, *Tdgf1*, *Gdf3*, *Nr0b1*, and *Fbxo15*) were not expressed or expressed at a lower level in PrE-primed cells and in c/eXEN. Differently, *Utf1*, *Tbx3*, and *Klf5* were expressed at higher levels in PDGFRα$^+$ cells than in epiblast-primed cells (Figure S4A) or e/cXEN (not expressed). By contrast, genes involved in extraembryonic specification were exclusively detected in PDGFRα$^+$ cells and e/cXEN. Remarkable differences were seen also for key pathway-associated genes. When compared with other analyzed lines, PDGFRα$^+$ cells expressed higher levels of LIF regulators, such as *Lifr* and *Il6st*, and Wnt-associated genes, such as *Lrp5/6* and *Dkk1*. Members of the fibroblast growth factor (FGF) signaling showed also a different pattern: *Fgf4* and *Fgfr1* levels were higher in epiblast-primed cells, while *Fgf3*, *Fgfr2*, and *Fgfr4* were exclusively expressed in PDGFRα$^+$ cells (Figure S4A).

Next, we focused on these subpopulations and performed gene set enrichment analysis (GSEA) of the PANTHER (protein analysis through evolutionary relationships) biological process and KEGG (Kyoto encyclopedia of genes and genomes) pathways. Genes upregulated in the PDGFRα$^+$ subpopulations were associated with metabolic processes and with lysosome, glutathione metabolism, and glycosphingolipid biosynthesis pathways (Tables S3 and S4). Epiblast-primed cells were enriched for terms associated with the cell cycle, focal adhesion, WNT and hedgehog signaling, cancer, developmental processes, and mesoderm/ectoderm development (Table S5). Remarkably, several terms enriched in the PDGFRα$^+$ subpopulations (highlighted in yellow in Tables S3 and S4) have been reported for 2i/L ESCs, while many terms enriched in the epiblast-primed subpopulation (highlighted in yellow in Table S5) have been reported for FBS/L ESCs when comparing the naive with the primed state of pluripotency (Marks et al., 2012).

We also performed unsupervised hierarchical clustering and principal component analysis (PCA) to visualize the relationship of the three subpopulations with published

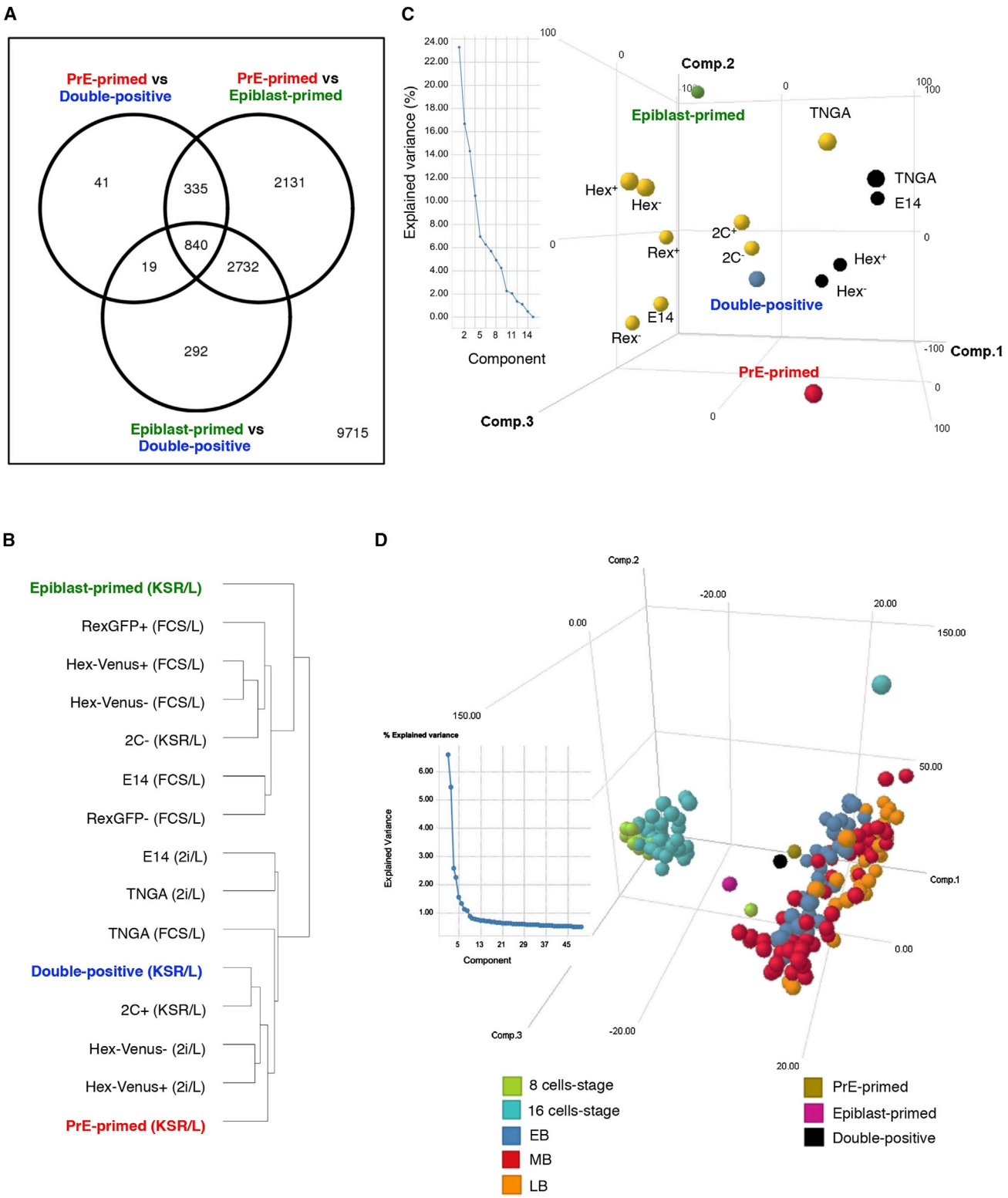


Figure 6. RNA-Seq Analysis and Comparison with In Vitro ESC Lines and In Vivo Single Cells

(A) Differentially expressed genes between sorted subpopulations. Adjusted p value < 0.05 with fold change (log2) > 1 or < −1.

(B) Unsupervised hierarchical clustering with previously published cell lines.

(legend continued on next page)

transcriptomes: two-cell stage (Macfarlan et al., 2012) ($2c^{+/-}$, KSR/L), $Rex1^{+/-}$ sorted cells, E14 and TNGA ESCs (Marks et al., 2012), and $Hex^{+/-}$ sorted cells (FBS/L and 2i/L) (Morgani et al., 2013). Unexpectedly, as ESCs cultured in 2i/L do not contain the PDGFRα$^+$ subpopulations, these analyses showed that the PDGFRα$^+$ subpopulations clustered more closely with naive than with other ESC lines (Figures 6B and 6C). Heatmap comparison of core/naive pluripotency and key pathway-associated genes with previously published transcriptomes (Figure S5) demonstrated that PDGFRα$^+$ subpopulations have a unique transcriptome. PDGFRα$^+$ cells, differently from the previously described *Hex*$^+$ cells, have a pronounced PrE-primed signature, while still retaining some pluripotency-related genes (*Oct4*, *Sall4*, *Utf1*, *Tbx3*, *Tfcp2l1*, and *Klf5*). PDGFRα$^+$ cells expressed *Dkk1* in an exclusive manner and had higher levels of LIF regulators (*Lifr* and *Il6st*) and FGF signaling members (*Fgf3/10*, *Fgfr2/3/4*).

As major differences could be detected between epiblast- and PrE-primed cells (Figures 6A and S4B), we assessed their relationship with single cells obtained from 8-cell-stage morula to late blastocyst (LB)-stage embryos (Deng et al., 2014). PCA analysis grouped a subset of early (EB) and mid (MB) blastocyst single cells with PrE-primed and with epiblast-primed cells (Figures 6D and S6A). GSEA of PrE/epiblast-primed subpopulations with the five most similar in vivo cells revealed upregulation of *Suz12* targets in the PrE cluster (Figure S6B) and DNA binding-related genes in the epiblast cluster (Figures S6C and S6D; Table S6). Thus, PDGFRα$^+$ cells have a unique expression profile and surprisingly show similarities with naive ESCs and with EB/MB cells in vivo.

Epigenetic Modifications and Signaling Involved in the Regulation of the PDGFRα$^+$ Subpopulations

To better understand the mechanisms and signaling governing this heterogeneity, we tested the effect of known epigenetic modifiers and small molecules. Considering the lower level of 5-mC in PrE-primed cells (Figures 4B and 4C), we added the DNA methylation inhibitor 5-azacytidine (5-AZA) to KSR/L. 5-AZA enhanced the frequency of PDGFRα$^+$ cells in a dose-dependent manner (Figure 7A). Likewise, the addition of dexamethasone, a glucocorticoid hormone involved in DNA demethylation and whose signaling interacts with the JAK/STAT pathway (Reddy et al., 2009), increased the PDGFRα$^+$ cell frequency (Figure 7B). The effects of 5-AZA and dexamethasone were combinatorial, resulting in ~3-fold increase in PDGFRα$^+$ cells (Figures 7B and S7A). Addition of trichostatin A (TSA), a histone deacetylase inhibitor, resulted in a decrease of PDGFRα$^+$ cells (Figure 7B), suggesting that histone acetylation negatively regulates the PrE-primed state.

As PDGFRα was shown to be necessary for eXEN derivation (Artus et al., 2010) and for conversion of ESCs into cXEN (Cho et al., 2012), we compared $Pdgfr\alpha^{H2B\text{-}GFP/+}$ (heterozygous) and $Pdgfr\alpha^{H2B\text{-}GFP/H2B\text{-}GFP}$ cells (a null knockin). The absence of the receptor did not alter the percentage of PDGFRα$^+$ cells (Figure 7C) or the transcript levels of pluripotent/extraembryonic genes in the PDGFRα$^+$ subpopulations (Figure S7B). Consistently, culture of ESCs with PDGF-AA did not significantly increase the percentage of PDGFRα$^+$ cells (Figure 7D).

As LIF supports the expansion of PrE in pre-implantation development (Morgani and Brickman, 2015), we tested if LIF was also necessary for the propagation of PDGFRα$^+$ cells. LIF withdrawal combined with the addition of a Janus kinase inhibitor (to block endogenous LIF), inhibited PDGFRα$^+$ cell expansion (Figures 7D and 7E). FGF signaling regulates the segregation of the PrE layer (Yamanaka et al., 2010) by phosphorylation of extracellular-signal-regulated kinase. The simultaneous inhibition of Gsk3β and MEK kinases in 2i/L resulted in the disappearance of PDGFRα$^+$ cells (Figure 7D); this effect was mediated by Mek (PD0325901), and not by the Gsk3β inhibitor, as shown by FACS for PDGFRα and OCT4/GATA4 (Figures S7C and S7D). This confirms the requirement of FGF also for the fluctuation of PDGFRα$^+$ cells.

DISCUSSION

During development, PDGFRα has an early, relatively weak but well visible expression from morula stage onward until it becomes stronger in PrE-fated cells at ~E3.75 (64 cells) (Artus et al., 2011; Grabarek et al., 2012; Plusa et al., 2008). In this study, we investigated its presence in undifferentiated ESCs, further dissecting their known heterogeneity. By taking advantage of the endogenous expression of PECAM1 and PDGFRα, we defined three different subpopulations that were further characterized (Figures 1 and 2): PECAM1$^+$/PDGFRα$^-$ (epiblast-primed), PECAM1$^+$/PDGFRα$^+$ (double-positive) and PECAM1$^-$/PDGFRα$^+$ (PrE-primed) cells. PrE-primed cells have a distinct molecular identity, as they co-express OCT4, GATA4, GATA6, and SOX17, which differs from epiblast-primed cells, which co-express OCT4, NANOG, and SOX2. Double-positive

(C) PCA analysis and explained variance with previously published cell lines. Cell lines with black dots were culture in 2i/L; cell lines with yellow dots were cultured with FBS/L, with the exception of $2C^{+/-}$, which were cultured in KSR/L.
(D) PCA analysis and explained variance with in vivo single cells from early embryonic stages.
See also Figures S4–S6.

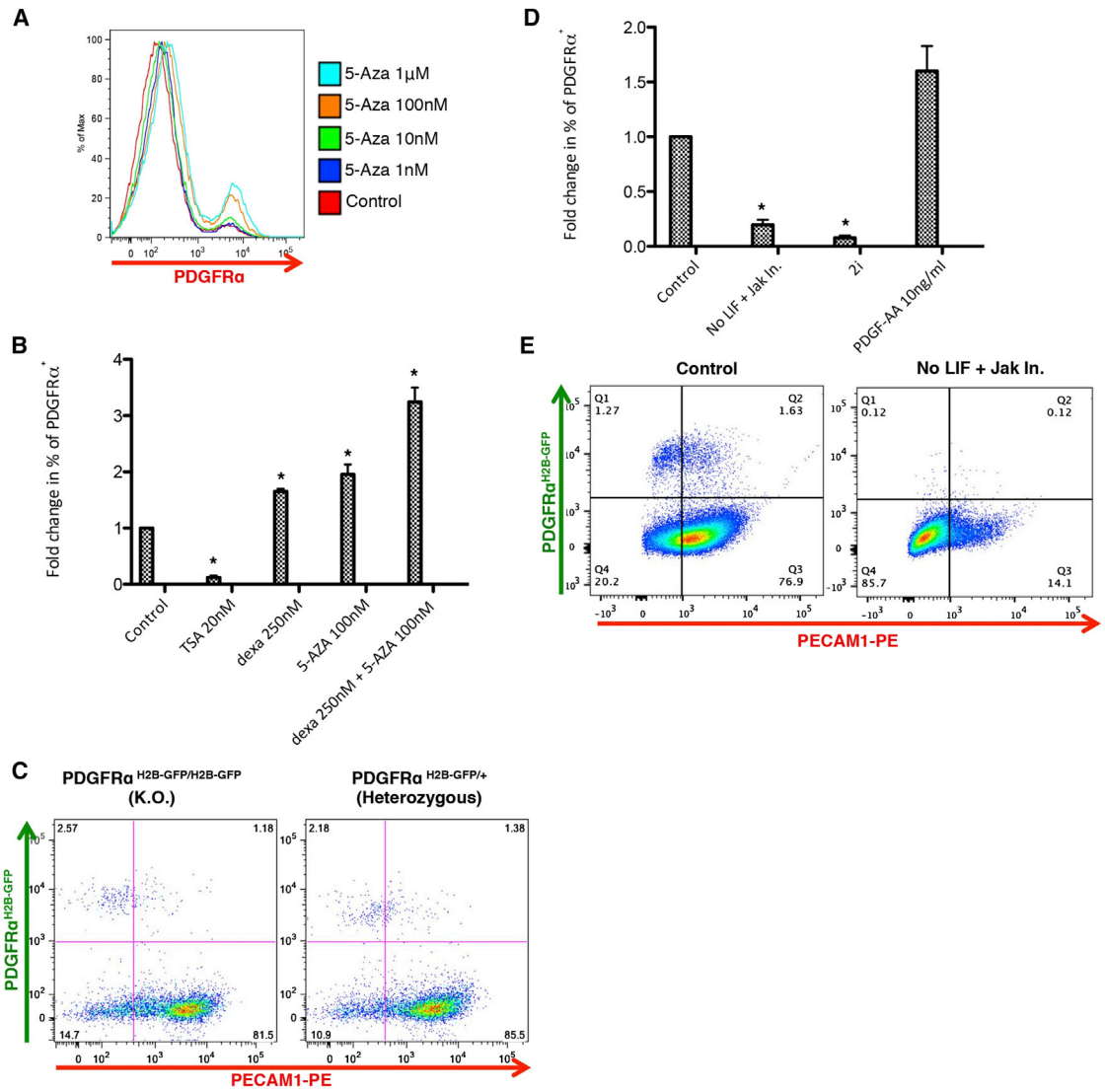


Figure 7. Epigenetic Modifications and Signaling Involved in the Regulation of PDGFRα+ Cells

(A) Dose response to 5-AZA treatment. Histograms show the percentage of PDGFRα+ cells in response to an increasing concentration of 5-AZA, n = 3.

(B) Fold change in percentage of PDGFRα+ cells after 72 hr of treatment under the indicated culture conditions for three independent experiments, *p < 0.05 by one-way ANOVA with subsequent Tukey honest significant difference(HSD) test.

(C) Representative FACS analysis for PDGFRα and PECAM1 in PDGFRα null and heterozygous ESC lines, n = 3. The gating strategy was based on isotype controls.

(D) Fold change in percentage of PDGFRα+ cells after 72 hr of treatment under the indicated culture conditions for three independent experiments, *p < 0.01 by one-way ANOVA with subsequent Tukey HSD test.

(E) Representative FACS analysis for PDGFRα and PECAM1 with LIF (left plot) or without LIF and with Jak Inhibitor (right plot), n = 3. The gating strategy was based on isotype controls.

See also Figure S7.

cells appear to be an intermediate between epiblast- and PrE-primed cells. In line with this, we also identified, at the single-cell level, cells co-expressing OCT4, GATA4, and NANOG. This is reminiscent of the simultaneous expression of epiblast and extraembryonic determinants in early pre-implantation development (Guo et al., 2010).

Although PrE-biased cells have already been described as *Hex*+ (Canham et al., 2010; Morgani et al., 2013), PDGFRα+ cells have a more pronounced PrE phenotype and a lower

expression of epiblast determinants, which are still retained in *Hex*$^+$ cells (Figure S5). Moreover, our model does not rely on signal amplification and on the use of reporter lines (Canham et al., 2010; Morgani et al., 2013), allowing the separation of these subpopulations in every ESC line of interest.

In vitro features of the three subpopulations appear to be drastically divergent in terms of self-renewal and differentiation capacity (Figure 3). Epiblast-primed cells and double-positive cells but not PrE-primed cells could re-establish the initial heterogeneity. When addressing their differentiation potential, PrE-primed cells efficiently generated XEN-like cells but not embryonic or presumptive trophoblast types diversely from epiblast-primed subpopulation.

PDGFRα$^+$ cells have a distinct epigenetic state, characterized by a lower level of DNA methylation/hydroxymethylation and by a different pattern of epigenetic regulators (Figure 4), in line with the notion that extraembryonic tissues (Sakaue et al., 2010) and their stem cell models (Rugg-Gunn et al., 2010) are hypomethylated.

The distinct epigenetic and molecular profile of PDGFRα$^+$ subpopulations was confirmed also by their developmental potential (Figure 5 and Table S1). The double-positive cells could still colonize the epiblast while PrE-primed cells exclusively contributed to PrE derivatives. Again, these in vivo experiments confirmed that PDGFRα$^+$ cells closely represent the PrE precursors.

The comparative transcriptome analysis with epiblast-primed cells and with e/cXEN showed that PDGFRα$^+$ subpopulations differentially express genes associated with core/naive pluripotency, and with JAK-STAT, WNT, and FGF signaling pathways (Figure S4). Unexpectedly, PCA, hierarchical clustering, and GSEA with previously available datasets, revealed that globally PDGFRα$^+$ cells resemble more naive ESCs (Figures 6B and 6C; Tables S3, S4, and S5). When compared with single cells obtained from early embryos, PrE-primed cells, as their epiblast counterpart, clustered with cells from the EB-MB stage (~E3.5–E4.0), further demonstrating that PDGFRα$^+$ steady states mirror the pre-implantation developmental window (Figures 6D and S6A).

The mechanisms involved in the regulation of the heterogeneity in vitro (Figure 7) confirmed previous studies in early development. The percentage of PDGFRα$^+$ cells was influenced by: JAK/STAT signaling, shown to support the expansion of PrE in pre-implantation development (Morgani and Brickman, 2015); FGF signaling, known to control the segregation of PrE and epiblast in the ICM (Yamanaka et al., 2010) and amount of DNA methylation, is a dispensable mechanism for the growth of extraembryonic lineages (Sakaue et al., 2010). By contrast, absence of PDGFRα, necessary for the derivation of eXEN (Artus et al., 2010) and cXEN (Cho et al., 2012), did not alter the abundance of PDGFRα$^+$ cells in vitro. Together, these results confirm that PDGFRα$^+$ cells are the in vitro equivalent of PrE precursors.

This model, which relies on the endogenous heterogeneous expression of PDGFRα, should facilitate and enable studies to gain insights in the factors regulating the early segregation of these different cell types within the ICM and to unravel the mechanisms involved in the different imprinting of embryonic and extraembryonic tissues (Hudson et al., 2010). Future studies are needed to determine whether PrE-primed cells recapitulate the imprinting associated with extraembryonic tissues (i.e., paternal imprinting of X chromosome) and whether a similar PrE-primed state is also present in human ESC cultures.

EXPERIMENTAL PROCEDURES

Cell Culture

Undifferentiated ESCs were maintained feeder free on gelatin (EmbryoMax 0.1% gelatin solution, ES-006-B; Millipore)-coated plates, in knockout DMEM (10829-018; Gibco), 20% knockout serum replacement (KSR, 10828-028; Gibco), 2 mM L-glutamine (25030-024; Gibco), 1x minimal essential medium nonessential amino acids (11140-035; Gibco), 1× penicillin-streptomycin (15140-122; Gibco), 100 μM β-mercaptoethanol (31350; Gibco) and 1,000 U/mL recombinant LIF (ESG1107; Chemicon International).

qRT-PCR Analysis

For RNA isolation, the RNeasy Mini-kit/Micro-kit (74104 and 74004; QIAGEN) was used. DNase treatment was achieved using the Turbo DNase kit (1907, Ambion). cDNA synthesis was done with 1 μg of RNA with the Superscript III First-Strand synthesis system (18080-051; Invitrogen). Real-time PCR was analyzed with the SYBR Green Platinum qPCR Supermix-UDG (11733-046; Invitrogen) on a ViiA 7 Real-Time PCR System (Applied Biosystems). Expression was normalized to *β-Actin*. Primer sequences are listed in the Supplemental Information.

Flow Cytometry and Cell Sorting

Single-cell suspensions of ESCs were obtained by dissociating with cell-dissociation buffer (Invitrogen) at 37°C for 20 min. Cells were washed twice with PBS and incubated with conjugated primary antibodies for 30 min on ice in the dark. Cells were washed once with PBS and resuspended for FACS analysis in PBS + 5% FBS. Flow cytometry was performed at the KU Leuven Flow Cytometry Facility using an FACS AriaIII (Becton Dickinson) or an FACS Canto (Becton Dickinson) for analysis. Intracellular staining was performed with the Foxp3/Transcription Factor Staining Buffer Set kit (00-5523-00; Ebioscience), following the manufacturer's protocol. Antibodies are listed in the Supplemental Information.

Differentiation Assays

Upon sorting of the different subpopulations, specific differentiations were performed as described by Sancho-Bru et al. (2011) for

mesendodermal differentiation; by Ying et al. (2003b) for neural differentiation and by Morgani et al. (2013) for TSC medium.

Culture Test

For the experiments described in Figures 7 and S7, 3×10^3 ESCs were sorted and plated in a 6-well plate in ESC medium under the following conditions: (1) no LIF and 1 μM InSolution JAK Inhibitor I (420097; Calbiochem); (2) 2i, 1 μM PD0325901 and 3 μM CHIR99021 (Axon Medchem); (3) 10 ng/mL PDGF-AA (315-18; Peprotech); (4) 20 nM Trichostatin A (TSA; T8552; Sigma); (5) 250 nM dexamethasone (D2915; Sigma); (6) 1 nM to 1 μM 5-AZA (A3656; Sigma); (7) 250 nM dexamethasone (Sigma) and 100 nM 5-AZA.

Immunoblotting

Sorted ESCs were lysed in RIPA buffer (R0278; Sigma) containing complete protease-inhibitor cocktail (04693116001; Roche) for 1 hr at 4°C. Protein concentrations of various samples were quantified using the Pierce BCA protein assay kit (23225; Thermo Scientific) following the manufacturer's instructions. To each protein sample, 1 volume of Bio-Rad loading buffer (161-0747, Bio-Rad) and β-mercaptoethanol (at 20:1, Sigma) was added. The samples were heated at 95°C for 10 min, followed by centrifugation at 13,000 × *g* for 10 min. Thirty micrograms of each protein sample was loaded in each lane of a 10% gradient Mini-PROTEAN TGXTM Precast gel (Bio-Rad) and electrophoresed. The resolved proteins were then transferred to Whatman Protan nitrocellulose membrane (Z613630; Sigma). Following blocking with 5% nonfat milk for 1 hr, membranes were incubated at 4°C overnight with primary antibodies. The following day, the membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies against rat, rabbit, and mouse IgG (Dako). Immunoreactive bands were visualized using Super Signal West Pico chemiluminescent substrate (34087; Thermo Scientific), and signals were detected using a ChemiDoc XRS$^+$ System (Bio-Rad). Antibodies are listed in the Supplemental Information.

Bisulfite Sequencing

Extraction of the genomic DNA isolated from FACS-sorted ESCs was done with the EpiTect Bisulfite kit (59104; QIAGEN). The primer sequences and PCR conditions for amplification of IAP sequences were as described (Lane et al., 2003). The PCR products were cloned using the pGEM-T Easy Vector System I (A1360; Promega). At least 15 colonies for each sample were sequenced and analyzed using Quma software (http://quma.cdb.riken.jp/).

5HmC/5mC Dot Blot Assay

FACS-sorted ESC subpopulations of genomic DNA was extracted with the PureLink Genomic DNA Mini kit (K182001; Invitrogen). Two-fold serial dilutions were made by mixing DNA and Tris-EDTA in 96-well plates. Twenty microliters of 1 M NaOH/25 mM EDTA was added to each well, the plate sealed, and heated at 95°C for 10 min. Subsequently, plates were cooled on ice and 50 μL of ice-cold 2 M ammonium acetate (pH 7.0) was added to each well, the plates were incubated on ice for 10 min. Subsequently, the denatured DNA was loaded on the nitrocellulose membrane (Bio-Rad), which was washed with 500 μL of 0.4 M NaOH, and rinsed with water. The membrane was air dried for 5–10 min and placed under UV (at 120,000 μJ/cm^2). The membrane was blocked with 5% nonfat milk in TBST (Tris-buffered saline and Tween 20) for 1 hr, and then incubated with antibodies against 5hmC/5mC O/N at 4°C. The membrane was washed with TBST for 10 min four times, and incubated with HRP-conjugated secondary antibodies (1: 5.000) at room temperature for 1 hr. The membrane was washed with TBST for 10 min four times, incubated with Enhanced ChemiLuminescence (ECL solution, Thermo Scientific) and developed using Chemidoc (Bio-Rad).

Statistical Analysis

p Values in the qRT-PCR analysis for pairwise differential expression against the epiblast-primed subpopulation were computed using Student's two-tailed t test. Experiments including three or more samples/treatment were subjected to one-way ANOVA with subsequent Tukey honest significant difference testing to establish significant changes between any two means.

Blastocyst Injections

Blastocyst injection studies were approved by the ethical committee for use of animals in research from KU Leuven (Belgium). C57BL/6 mouse ESCs were labeled with eGFP by lentiviral transduction. The eGFP transcription was under the control of elongation factor-1alpha promoter. Following culture in KSR/L, cells were dissociated and different subpopulations were sorted based on PDGFRα and PECAM1 labeling. Sorted cells (6–8 cells) were immediately injected in the blastocoel cavity of CD1 blastocysts. The embryos were transferred the same day to the uterus of pseudopregnant CD1 female mice. Post-implantation embryos were collected at E6.5 from pseudopregnant mice 3.5 days after embryo transfer. Intact post-implantation conceptuses were isolated from decidua, fixed with 4% paraformaldeyde, and immediately imaged using a SteREO Discovery V12 microscope (Zeiss) to determine chimerism.

ACCESSION NUMBERS

The gene expression data reported in this paper has been deposited at the GEO repository with accession number GEO: GSE65884.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2016.12.010.

AUTHOR CONTRIBUTIONS

C.M.V. and A.L.N. designed the project and experiments. A.L.N, G.M.F. performed most of the experiments. I.P. helped with immunofluorescences and W.B., S.M.C.d.S.L. helped with the interpretation of the results. X.L.A., P.G.C., and K.-P.K provided materials. A.d.J. and F.L. performed differentiation experiments and helped with the revision, A.Z., R.K., and V.A.E. embedded/analyzed chimeric embryos. D.S.C. and W.S.H. performed RNA-seq analysis.

A.L.N. and C.M.V. wrote the manuscript. F.L. and C.M.V. contributed equally. All the authors read the manuscript.

ACKNOWLEDGMENTS

We thank the late Vik Van Duppen for FACS experiments, Rob Van Rossom and the KU LEUVEN FACS CORE for sorting, Zhiyong Zhang and Liesbeth Vermeire from InfraMouse for chimera production, and Kristel Eggermont for help with image acquisition/processing. We thank Dr. Hadjantonakis and Dr. Niakan for the FGF4/PDGFRα ESC lines and Dr. Morrison for the Sox17$^{GFP/+}$ line. We thank Dr. Kian Koh and Joris Vande Velde for dot blots. The work was supported by grants obtained from FWO (G.0832) and KU Leuven (EIW-B4855-EF/05/11 and ETH-C1900-PF to C.M.V.; EME-C2161-GOA/11/012 to C.M.V./A.Z.; C14/16/078 to F.LL.), Hercules Foundation (ZW09/03 to A.Z.) and by the BELSPO-IUAP-DEVREPAIR grant (to C.M.V./S.M.C.d.S.L./A.Z.).

Received: October 22, 2015
Revised: December 9, 2016
Accepted: December 12, 2016
Published: January 12, 2017

REFERENCES

Alder, O., Lavial, F., Helness, A., Brookes, E., Pinho, S., Chandrashekran, A., Arnaud, P., Pombo, A., O'Neill, L., and Azuara, V. (2010). Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment. Development *137*, 2483–2492.

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat. Rev. Mol. Cell Biol. *10*, 91–103.

Artus, J., Panthier, J.J., and Hadjantonakis, A.K. (2010). A role for PDGF signaling in expansion of the extra-embryonic endoderm lineage of the mouse blastocyst. Development *137*, 3361–3372.

Artus, J., Piliszek, A., and Hadjantonakis, A.K. (2011). The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. Dev. Biol. *350*, 393–404.

Beddington, R.S., and Robertson, E.J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development *105*, 733–737.

Bessonnard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G., and Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. Development *141*, 3637–3648.

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. Nat. Cell Biol. *16*, 516–528.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Bryja, V., Bonilla, S., Cajanek, L., Parish, C.L., Schwartz, C.M., Luo, Y., Rao, M.S., and Arenas, E. (2006). An efficient method for the derivation of mouse embryonic stem cells. Stem Cells *24*, 844–849.

Canham, M.A., Sharov, A.A., Ko, M.S., and Brickman, J.M. (2010). Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biol. *8*, e1000379.

Cho, L.T., Wamaitha, S.E., Tsai, I.J., Artus, J., Sherwood, R.I., Pedersen, R.A., Hadjantonakis, A.K., and Niakan, K.K. (2012). Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. Development *139*, 2866–2877.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

Dietrich, J.E., and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. Development *134*, 4219–4231.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Grabarek, J.B., Zyzynska, K., Saiz, N., Piliszek, A., Frankenberg, S., Nichols, J., Hadjantonakis, A.K., and Plusa, B. (2012). Differential plasticity of epiblast and primitive endoderm precursors within the ICM of the early mouse embryo. Development *139*, 129–139.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev. Cell *18*, 675–685.

Hamilton, T.G., Klinghoffer, R.A., Corrin, P.D., and Soriano, P. (2003). Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms. Mol. Cell. Biol. *23*, 4013–4025.

Hayashi, K., Lopes, S.M., Tang, F., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell *3*, 391–401.

Hayashi, Y., Furue, M.K., Tanaka, S., Hirose, M., Wakisaka, N., Danno, H., Ohnuma, K., Oeda, S., Aihara, Y., Shiota, K., et al. (2010). BMP4 induction of trophoblast from mouse embryonic stem cells in defined culture conditions on laminin. In Vitro Cell. Dev. Biol. Anim. *46*, 416–430.

Hudson, Q.J., Kulinski, T.M., Huetter, S.P., and Barlow, D.P. (2010). Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. Heredity *105*, 45–56.

Kunath, T., Arnaud, D., Uy, G.D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R.L., Avner, P., and Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development *132*, 1649–1661.

Kwon, G.S., Viotti, M., and Hadjantonakis, A.K. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. Dev. Cell *15*, 509–520.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis *35*, 88–93.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Lo Nigro, A., Geraerts, M., Notelaers, T., Roobrouck, V.D., Muijtjens, M., Eggermont, K., Subramanian, K., Ulloa-Montoya, F., Park, Y., Owens, J., et al. (2012). MAPC culture conditions support the derivation of cells with nascent hypoblast features from bone marrow and blastocysts. J. Mol. Cell Biol. *4*, 423–426.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57–63.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., et al. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. USA *78*, 7634–7638.

Morgani, S.M., and Brickman, J.M. (2015). LIF supports primitive endoderm expansion during pre-implantation development. Development *142*, 3488–3499.

Morgani, S.M., Canham, M.A., Nichols, J., Sharov, A.A., Migueles, R.P., Ko, M.S., and Brickman, J.M. (2013). Totipotent embryonic stem cells arise in ground-state culture conditions. Cell Rep. *3*, 1945–1957.

Niakan, K.K., Schrode, N., Cho, L.T., and Hadjantonakis, A.K. (2013). Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. Nat. Protoc. *8*, 1028–1041.

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oles, A.K., Arauzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.K., et al. (2014). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat. Cell Biol. *16*, 27–37.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development *135*, 3081–3091.

Reddy, T.E., Pauli, F., Sprouse, R.O., Neff, N.F., Newberry, K.M., Garabedian, M.J., and Myers, R.M. (2009). Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. Genome Res. *19*, 2163–2171.

Rossant, J., Sanford, J.P., Chapman, V.M., and Andrews, G.K. (1986). Undermethylation of structural gene sequences in extraembryonic lineages of the mouse. Dev. Biol. *117*, 567–573.

Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc. Natl. Acad. Sci. USA *107*, 10783–10790.

Sakaue, M., Ohta, H., Kumaki, Y., Oda, M., Sakaide, Y., Matsuoka, C., Yamagiwa, A., Niwa, H., Wakayama, T., and Okano, M. (2010). DNA methylation is dispensable for the growth and survival of the extraembryonic lineages. Curr. Biol. *20*, 1452–1457.

Sancho-Bru, P., Roelandt, P., Narain, N., Pauwelyn, K., Notelaers, T., Shimizu, T., Ott, M., and Verfaillie, C. (2011). Directed differentiation of murine-induced pluripotent stem cells to functional hepatocyte-like cells. J. Hepatol. *54*, 98–107.

Schrode, N., Saiz, N., Di Talia, S., and Hadjantonakis, A.K. (2014). GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. Dev. Cell *29*, 454–467.

Wicklow, E., Blij, S., Frum, T., Hirate, Y., Lang, R.A., Sasaki, H., and Ralston, A. (2014). HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. PLoS Genet. *10*, e1004618.

Yamanaka, Y., Lanner, F., and Rossant, J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development *137*, 715–724.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003a). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281–292.

Ying, Q.L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003b). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat. Biotechnol. *21*, 183–186.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Stem Cell Reports, Volume *8*

Supplemental Information

PDGFRα^+ Cells in Embryonic Stem Cell Cultures Represent the In Vitro Equivalent of the Pre-implantation Primitive Endoderm Precursors

Antonio Lo Nigro, Anchel de Jaime-Soguero, Rita Khoueiry, Dong Seong Cho, Giorgia Maria Ferlazzo, Ilaria Perini, Vanesa Abon Escalona, Xabier Lopez Aranguren, Susana M. Chuva de Sousa Lopes, Kian Peng Koh, Pier Giulio Conaldi, Wei-Shou Hu, An Zwijsen, Frederic Lluis, and Catherine M. Verfaillie

SUPPLEMENTAL FIGURES

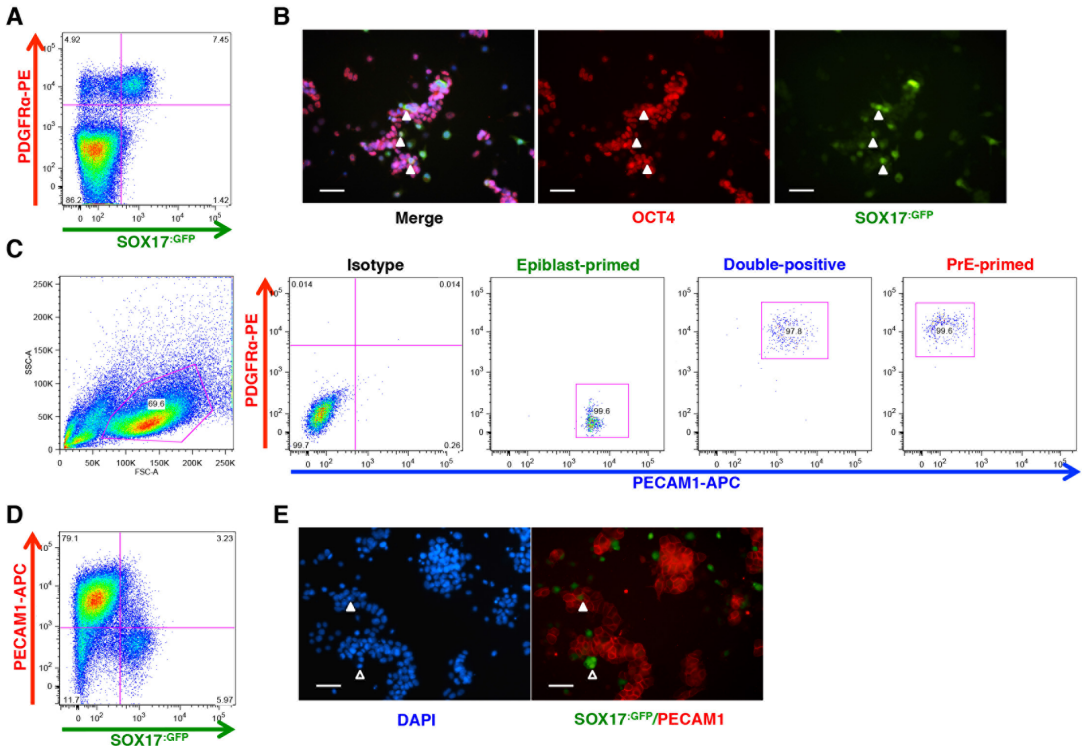


Figure S1. Co-expression of Sox17 with PDGFRα, Oct4 and PECAM1; related to Figure 1.

(A) Representative FACS analysis on *Sox17*$^{GFP/+}$ line for the expression of PDGFRα, n=3. Gating strategy was based on isotype controls.

(B) Immunostaining analysis for OCT4 on *Sox17*$^{GFP/+}$ line. Arrows indicate cells coexpressing OCT4 and SOX17. Scale bar=50μm, n=3.

(C) FSC/SSC plot, isotype controls, gating strategies and sorting purities (relative to all experiments involving sorting).

(D) Representative FACS analysis on *Sox17*$^{GFP/+}$ line for the expression of PECAM1, n=3. Gating strategy was based on isotype controls.

(E) Immunostaining analysis on *Sox17*$^{GFP/+}$ for the expression of PECAM1. Full arrow indicates cells co-expressing PECAM1 and SOX17. Empty arrow indicates cells expressing only SOX17. Scale bar=50μm, n=3.

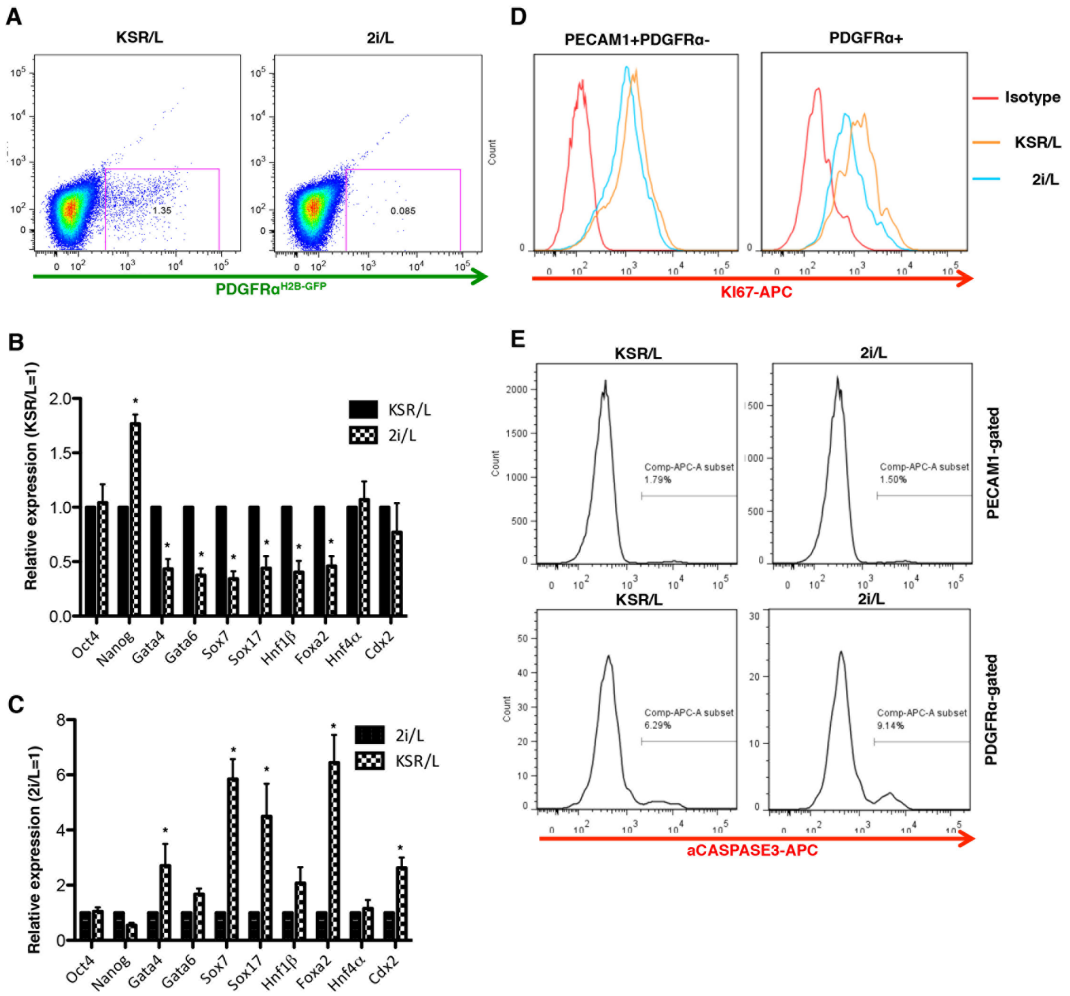


Figure S2. Effect of 2i on PDGFRα and PrE-related genes expression; related to Figure 1.

(A) Representative FACS analysis showing the percentage of *Pdgfrα*$^{H2B\text{-}GFP/+}$ cells cultured in KSR/L (left plot) or 2i/L (right plot), n=3.

(B) qRT-PCR analysis for embryonic and extraembryonic TFs on *Pdgfrα*$^{H2B\text{-}GFP/+}$ line (previously cultured in KSR/L) after 72 hours in 2i/L. Data are represented as Mean ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *p < 0.05, t test.

(C) qRT-PCR analysis for embryonic and extraembryonic TFs on *Pdgfrα*$^{H2B\text{-}GFP/+}$ line (previously cultured in 2i/L) after 72 hours in KSR/L. Data are represented as Mean ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *p < 0.05, t test.

(D) Representative FACS analysis for PDGFRα, PECAM1 and KI67 on R1 line, cultured in KSR/L or 2i/L. n=3.

(E) Representative FACS analysis for PDGFRα, PECAM1 and active Caspase3 on R1 line, cultured in KSR/L or 2i/L. n=3.

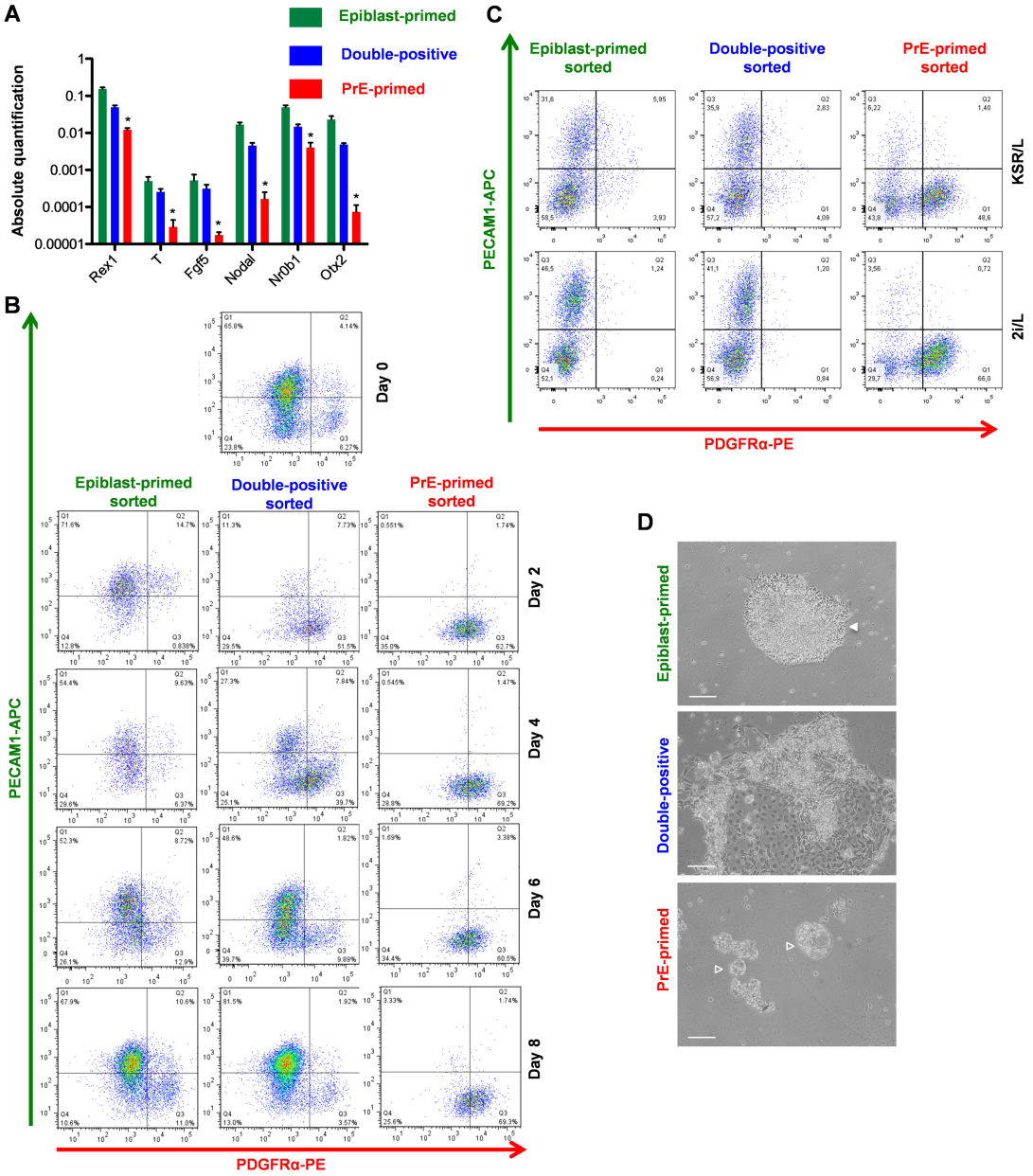


Figure S3. *In vitro* functional differences of the PrE-primed subpopulations; related to Figure 3.

(A) qRT-PCR analysis for post-implantation epiblast markers, after sorting. Data are represented as Mean ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *$p < 0.05$, t test.

(B) Representative time-course FACS analysis of the sorted subpopulations for PDGFRα and PECAM1 during re-establishment of the initial heterogeneity (E14 line). n=3. Gating strategy was based on isotype controls.

(C) Representative FACS analysis of the sorted subpopulations for PDGFRα and PECAM1 cultured in KSR/L or 2i/L for a week (E14 line). n=3. Gating strategy was based on isotype controls.

(D) Bright field pictures of sorted subpopulations cultured in neural-promoting condition. Full arrows indicate neuroectodermal precursors, while empty arrows indicate vacuolated structures. Scale bar=50μm.

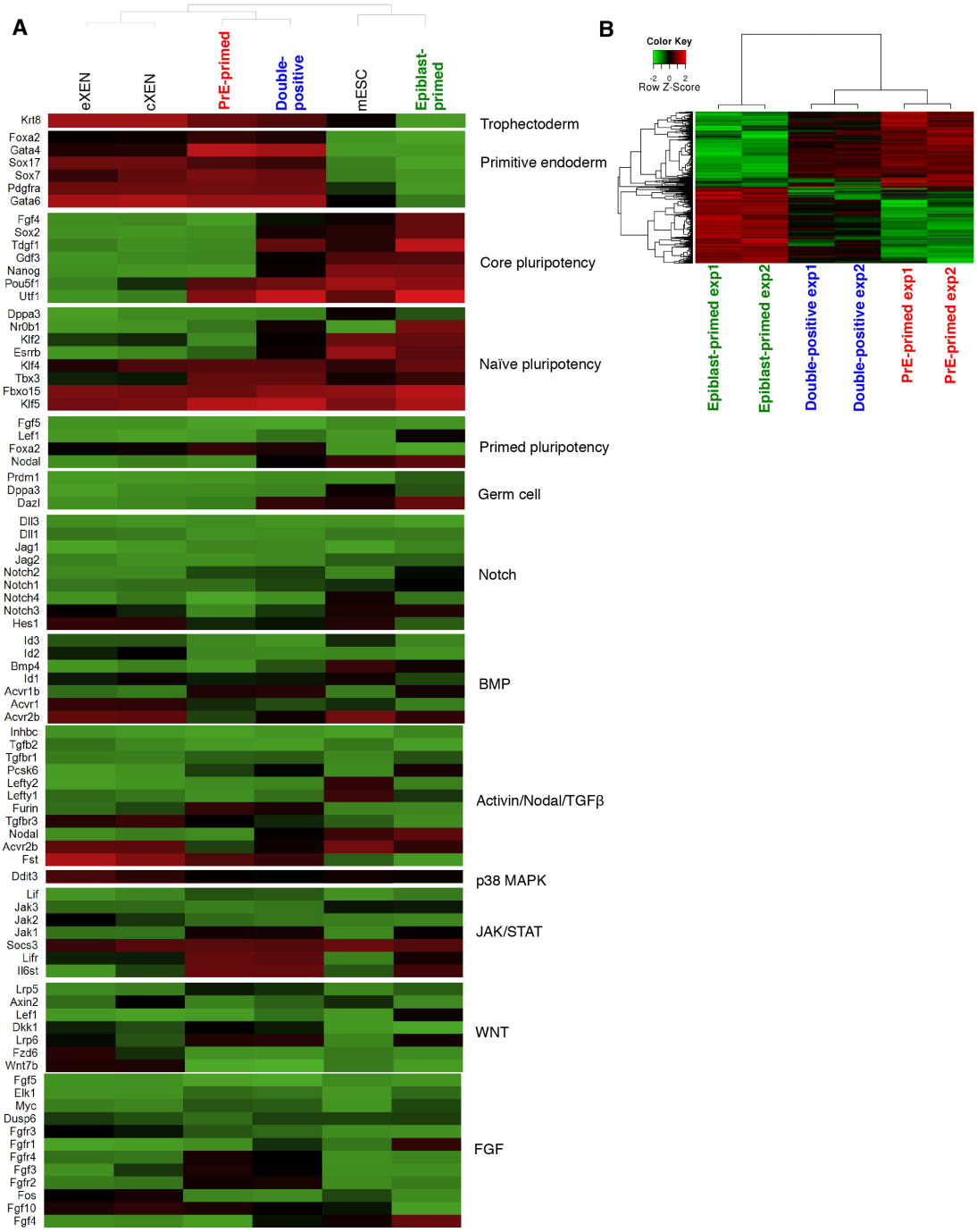


Figure S4. RNA-Seq analysis on sorted subpopulations and comparison to e/cXEN; related to Figure 6.

(A) Heat map of Naive/extraembryonic and pathway related genes on sorted subpopulations, eXEN, cXEN and unsorted ESC. Heat-map shows differentially expressed genes identified by pairwise comparison of sorted fractions. Genes are hierarchically clustered by average Euclidean distance.

(B) Heat map of sorted subpopulations, based on RNA-seq data. Heat-map shows differentially expressed genes identified by pairwise comparison of sorted fractions. Genes are hierarchically clustered by average Euclidean distance. Two biological replicates per sample are shown. Red represents upregulation while green downregulation.

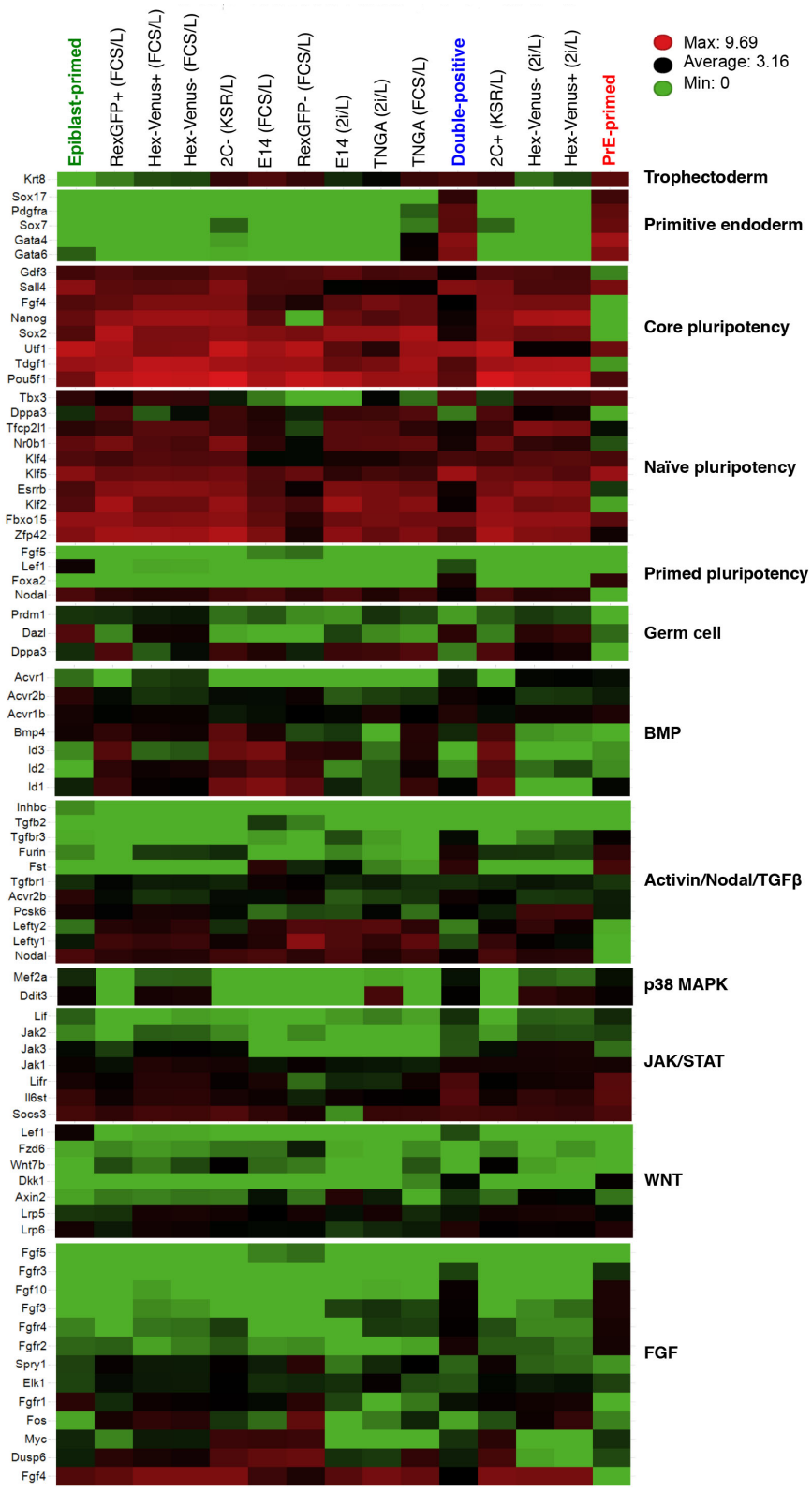


Figure S5. Heat-map comparison with other cell lines; related to Figure 6.

Heat map comparison of Naive/extraembryonic and pathway related genes on sorted subpopulations and previously published cell lines.

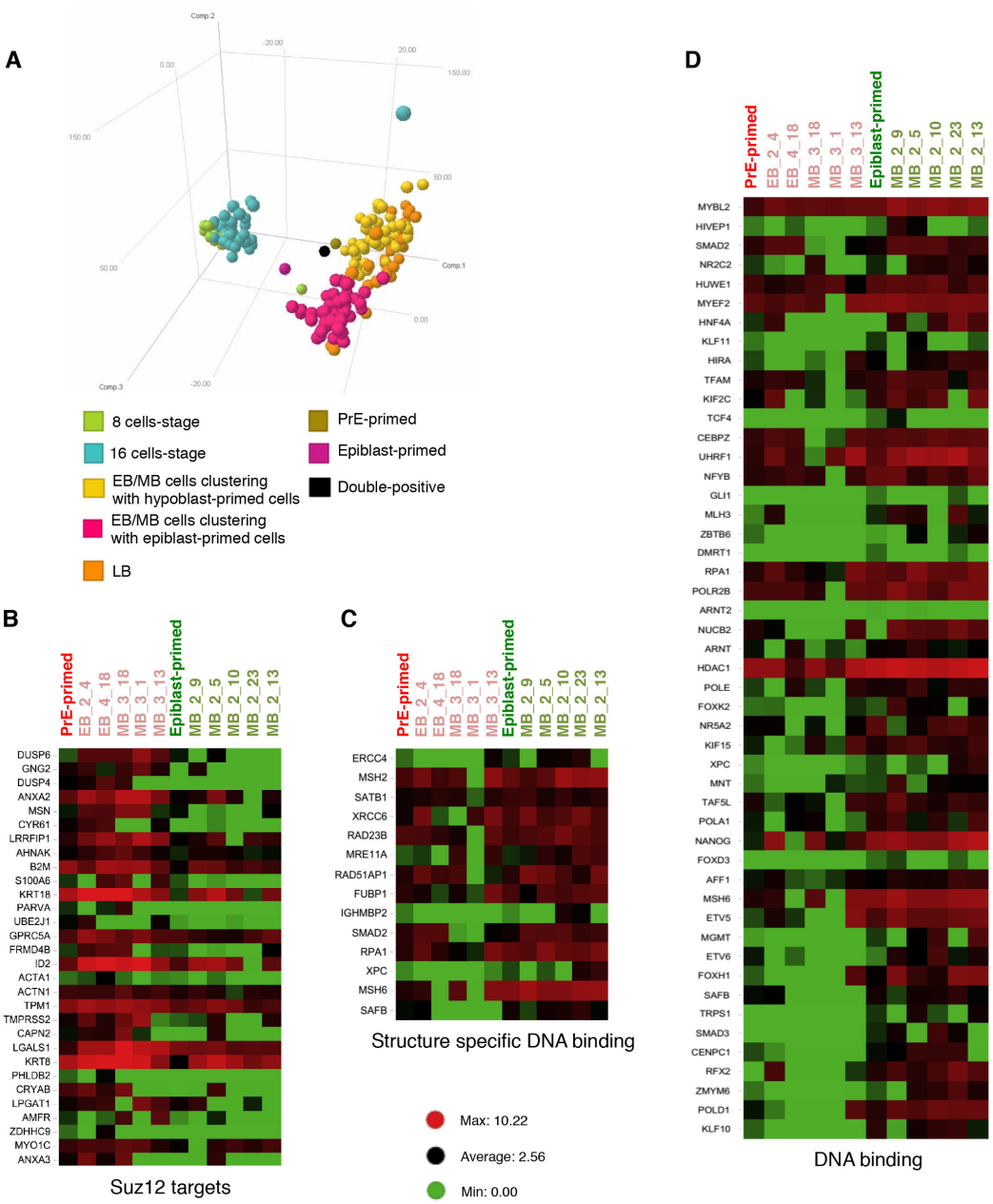


Figure S6. RNA-Seq comparison with *in vivo* cells; related to Figure 6.

(A) PCA analysis and explained variance with in vivo single cells showing EB/MB cells clustering with PrE-primed (in yellow) and epiblast-primed (in pink) subpopulations.

(B) Heat map of Suz12 targets, after GSEA of PrE- and epiblast-primed cells with their five most similar *in vivo* cells.

(C) Heat map of structure specific DNA binding-related genes, after GSEA of PrE- and epiblast-primed cells with their five most similar *in vivo* cells.

(D) Heat map of DNA binding-related genes, after GSEA of PrE- and epiblast-primed cells with their five most similar *in vivo* cells.

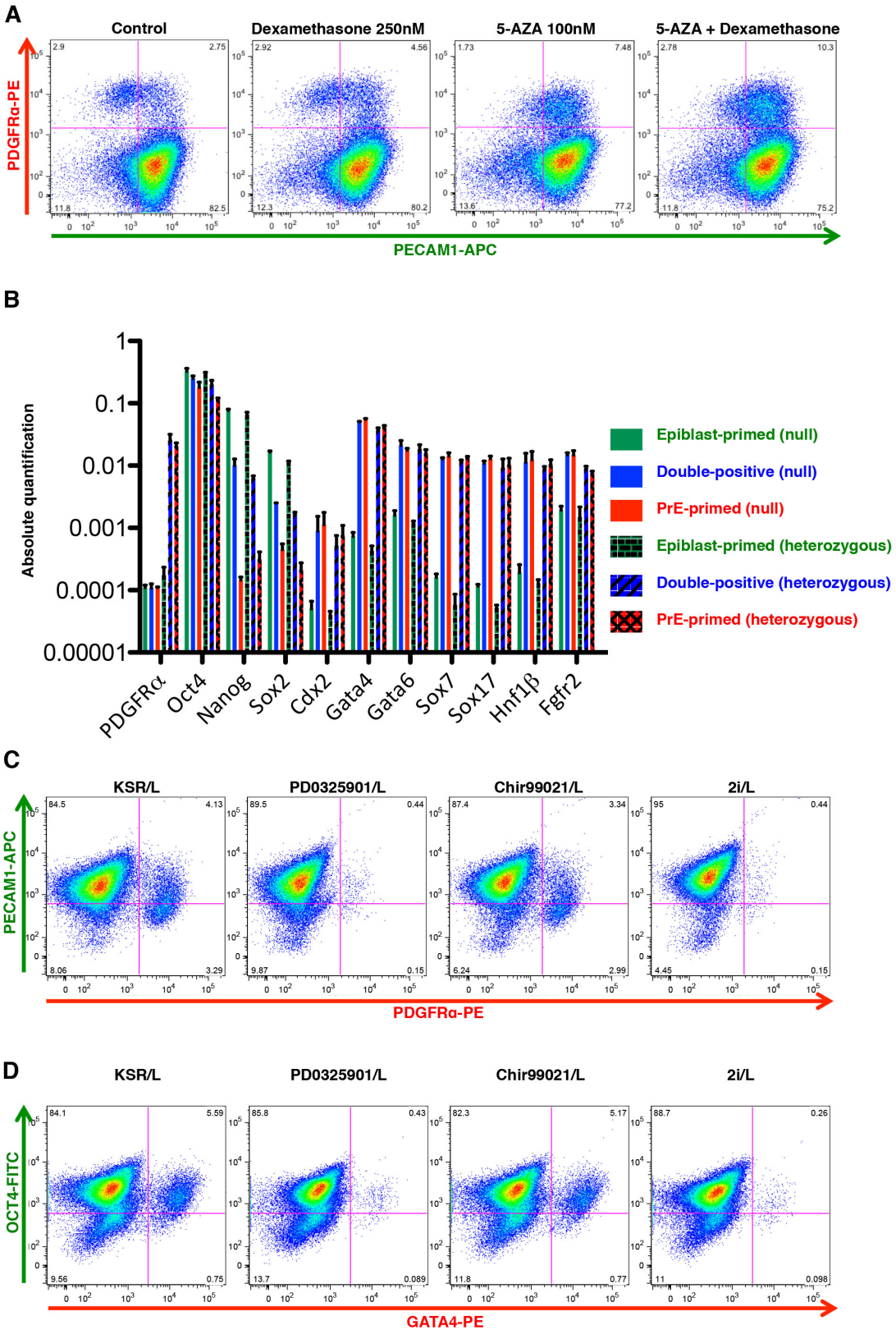


Figure S7. Epigenetic modifications and signaling involved in the regulation of the PrE-primed subpopulations; related to Figure 7.

(A) Representative FACS analysis for PDGFRα and PECAM1 on ESC cultured with Dexamethasone, 5-AZA and their combination, n=3. Gating strategy was based on isotype controls.

(B) qRT-PCR analysis for embryonic and extraembryonic markers in *Pdgfrα*$^{H2B\text{-}GFP/H2B\text{-}GFP}$ (null) and *Pdgfrα*$^{H2B\text{-}GFP/+}$ (heterozygous) ESC line. Data are represented as Mean ± SEM of each transcript from three independent experiments (normalized to *β-Actin*), *p < 0.05, t test.

(C) Representative FACS analysis for PDGFRα and PECAM1 on ESC cultured with MEK inhibitor, GSK3 inhibitor and 2i, n=3. Gating strategy was based on isotype controls.

(D) Representative intracellular FACS analysis for OCT4 and GATA4 on ESC cultured with MEK inhibitor, GSK3 inhibitor and 2i, n=3. Gating strategy was based on isotype controls.

SUPPLEMENTAL TABLE LEGENDS

Table S1: Summary of blastocyst injections.

Table S2: RNA sequencing on sorted samples. The excel file contains RNA-seq data of the tree different subpopulations

Table S3: GSEA on PrE-primed cells. The excel file contains gene set enrichment analysis for PrE-primed cells. Terms highlighted in yellow have been reported for 2i/L ESC (ground-state pluripotency).

Table S4: GSEA on double-positive cells. The excel file contains gene set enrichment analysis for double-positive cells. Terms highlighted in yellow have been reported for FBS/L ESC (primed pluripotency).

Table S5: GSEA on epiblast-primed cells. The excel file contains gene set enrichment analysis for epiblast-primed cells. Terms highlighted in yellow have been reported for FBS/L ESC (primed pluripotency).

Table S6: GSEA between *in vitro/in vivo* cells. The excel file contains gene set enrichment analysis for PrE/epiblast-primed cells and their five most similar cells from early/mid blastocyst stages.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Immunostaining cellular material

Cells were washed with PBS and fixed for 10 min with a 4% formaldehyde solution, permeabilised in 0.2% Triton-X-100 (Sigma) in PBS, incubated with a blocking solution (5% donkey serum in PBS) and stained overnight at 4°C with primary antibodies diluted in DAKO antibody diluent. Samples were incubated for 30 min at RT with the secondary antibodies in DAKO antibody diluent together with Hoechst 33258, diluted to 1:2000, for nuclear staining. In between the incubation steps, cells were washed with PBS containing 0.2% Triton-X-100. The immunostained cells were examined with a Nikon Eclipse Ti microscope using the 10x, 20x and 40x objectives, and using the Image pro plus software. Antibodies are listed in the dedicated table.

RNA-Sequencing

Library preparation. We extracted RNA from 10^5 sorted events for each subpopulation, using the Mirneasy RNA micro kit (QUIAGEN). RNA was amplified by using the Ovation RNA Amplification System V2 (NUGEN). cDNA profile was checked by running the samples on a BioAnalyzer 2100 instrument (Agilent Technologies, Santa Clara, USA) using a High Sensitivity DNA chip. cDNA quantification was done with the Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, USA) using the Qubit dsDNA HS Assay Kit. The concentrations ranged from 69.4 - 99.8 ng/µl.For all samples 1000 ng cDNA was sheared to 300 bp using the Covaris M220 Focused Ultrasonicator (SonoLab 7 Software) and screw cap micro tubes of 50 µl. Result was checked by running the sheared samples on a BioAnalyzer 2100 instrument on a High Sensitivity DNA chip. The concentration of the sheared samples was measured on the Qubit 2.0 Fluorometer. Measured concentrations were all between 17.0 and 19.8 ng/µl. Finally, library preparation was done following the manufacturer's NEBNext Ultra DNA Library preparation protocol (NEB#E7370L version 1.0, New England Biolabs) with the following minor modification: only halve of the adaptor ligated DNA fragments are subjected to a 10 cycle PCR in the final library amplification step. As input for the library preparation, 100 ng of sheared cDNA in a volume of 55.5 µl was used. Size selection was done following the recommendation of the

protocol with insert size 250 bp and total Library Size of 400 bp. The used barcoded adaptors (NEBNext Multiplex Oligos for Illumina, NEB#E7335L (Set A) + NEB#7500L (Set B)) were A001, A003, A005, A006, A012 and A019. Adaptors were checked with Illumina Experiment Manager. Final libraries were quantified using the Qubit High Sensitivity assay (Life Technologies, Carlsbad, USA). Size distribution and average length were determined by running the libraries on a BioAnalyzer 2100 using the DNA High Sensitivity chip. Because of the presence of primer dimer and over-amplification in library 2 and 4, these libraries were subjected to a Double Side Size Selection with SPRI Select Beads (Beckman Coulter). Concentration and BA profile were checked again. Molarity of each library was calculated from the concentration and the average insert length and ranged between 16.12 and 65.8 nM. Individual libraries were equimolar pooled to a 10 nM pool in a total volume of 60 µl. This pool was mixed with another 10 nM pool (60 µl) using different adaptors.

Sequencing and statistical data analysis. Sequencing was performed on 1/2 lane on an Illumina HiSeq 2000 using the 50 bp single read recipe at the Genomics Core (Leuven, Belgium).

Preprocessing. Low quality ends and adapter sequences were trimmed off from the Illumina reads with FastX 0.0.13 and Cutadapt 1.2.1 (http://hannonlab.cshl.edu/fastx_toolkit/index.html; Martin, 2011). Using FastX 0.0.13 and ShortRead 1.20.0, we filtered subsequently small reads (length < 35 bp), polyA-reads (more than 90% of the bases equal A), ambiguous reads (containing N) and low quality reads (more than 50% of the bases < Q25) (M. Morgan, 2009). With Bowtie2 v2.1.0 we identified and removed reads that mapped to the spiked-in PhiX (Langmead and Salzberg, 2012). The number of processed reads per sample then varied between 11,247,851 and 14,399,346.

Mapping. Processed reads were aligned with Tophat v2.0.8b to the reference genome of Mus musculus (GRCm38.73), as downloaded from the Genome Reference Consortium (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/; Trapnell et al., 2009). Default Tophat parameter settings were used, except for 'min-intron-length=50', 'max-intron-length=500,000', 'no-coverage-search' and 'read-realign-edit-dist=3'. Using Samtools 0.1.19, reads with a mapping quality smaller than 20 were removed from the alignments(Li et al., 2009).

Counting - Transcript coordinates were extracted from the GRC reference annotation with Gffread from the Cufflinks v2.1.1 suite and merged to gene coordinates with mergeBed from the Bedtools v2.17.0 toolkit (Quinlan and Hall, 2010; Trapnell et al., 2010). GC content and gene length were derived from the gene coordinates. The numbers of aligned reads per gene were summarized using HTSeq-count v0.5.4p3 with parameters 'm=union', 'stranded=no', 'a=0', 't=exon' and 'i=gene id' (http://www-huber.embl.de/users/anders/HTSeq). We removed 22,456 genes for which all samples had less than 1 count-per-million. As such, we continued with raw counts for 16,105 genes. Raw counts were further corrected within samples for GC-content and between samples using full quantile normalization, as implemented in the EDASeq 1.8.0 package from Bioconductor (Risso et al., 2011).

Identifying differential expression. With the EdgeR 3.4.0 package, a negative binomial generalized linear model (GLM) was fitted against the normalized counts (Robinson and Smyth, 2007). We did not use the normalized counts directly, but worked with offsets. Differential expression was tested for with a GLM likelihood ratio test, also implemented in the EdgeR package. The resulting p-values were corrected for multiple testing with Benjamini-Hochberg to control the false discovery rate. Statistical analysis was done at the Nucleomics Core (Leuven, Belgium).

RNA-seq analysis and comparison

RNA-seq datasets compared were downloaded from Gene Expression Omnibus (GEO) and Sequence Read Archive. RPKM values of single cells from mouse embryos at different preimplantation stage(Deng et al., 2014) was obtained from GSE45719. We downloaded RNA-seq data of 2C::tdTomato+, 2C::tdTomato-(Macfarlan et al., 2012), TNGA-2i, TNGA-serum, E14-serum, E14-2i, Rex1GFP-negative, and Rex1GFP-positive ES cells(Marks et al., 2012) from SRR385620, SRR385621, SRR064969, SRR064970, SRR064971, SRR064972, SRR392299, and SRR392300, respectively. Hex-Venus+ and Hex-Venus- ES cells in serum/LIF or 2i/LIF (Morgani et al., 2013) were obtained from GSE45719 and GSE45182, respectively. Illumina microarray data for mESCs, cXEN cells, and embryo-derived XEN cells (Cho et al., 2012) was obtained from GSE38477

They were aligned by Tophat2, and RPKM values are computed by Cufflink. Data across different publications was combined based on matching their ENSEMBL identifiers. Hierarchical clustering showed that data from each batch clustered independently, therefore empirical Bayes methods(Johnson et al., 2007) were used to eliminate these batch dependent effects and allow for data comparison.mPrincipal component analysis (PCA) and hierarchical clustering were performed using

OmicsOffice built in TIBCO Spotfire. Expression intensities were clustered by complete linkage method and their similarity was measured as Euclidean distance for hierarchical clustering.
Gene set enrichment analysis (GSEA) was done by GSEA software (Subramanian et al., 2005). The analysis was performed based on KEGG gene sets (http://www.broadinstitute.org/gsea/msigdb/collections.jsp#C2) downloaded from Broad Institute with 1000 permutations. GO and KEGG analysis between different sorted subpopulations was performed with DAVID (http://david.abcc.ncifcrf.gov/)

Table of primers used for this study

GENE	FORWARD	REVERSE	APPLICATION
Afp	CATGCTGCAAAGCTGACAA	CTTTGCAATGGATGCTCTCTT	qRT-PCR
B-Actin	TGTTACCAACTGGGACGACA	GGGGTGTTGAAGGTCTCAAA	qRT-PCR
Cdx2	AAGACAAATACCGGGTGGTG	CCAGCTCACTTTTCCTCCTG	qRT-PCR
Cxcr4	GCTCACCCTCATTACATACA	TAGAACTCAACAGGAGGCGG	qRT-PCR
Dnmt1	GGGTCTCGTTCAGAGCTG	GCAGGAATTCATGCAGTAAG	qRT-PCR
Dnmt3a	CCTGCAATGACCTCTCCATT	CAGGAGGCGGTAGAACTCAA	qRT-PCR
Dnmt3b	CCAAGGACACCAGGACGCGC	TCCGAGACCTGGTAGCCGGAA	qRT-PCR
Dnmt3l	CTGCTGACTGAGGATGACCA	GCTTGCTCCTGCTTCTGACT	qRT-PCR
Eomes	AGAACCGTGCCACAGACCAA	TGGTCACAGGTTGCTGGACA	qRT-PCR
Esrrb	AACCATTCAAGGCAACATCG	TTTGAGGCATTTCATGAATCGG	qRT-PCR
Fgf5	CGGACGGTGAACGACTACAC	CGTTGGAGAACCTCACTTGAC	qRT-PCR
Fgfr2	TCGATAAAGACAAACCCAAGGAG	AGATCAGACAGGTCCTTCTCTG	qRT-PCR
Foxa2	CGAGTTAAAGTATGCTGGGAG	TATGTGTTCATGCCATTCATCC	qRT-PCR
Gata3	CACAACGCAGAGCTAAGCAA	TTGTAGTTGGGGTGGTCCTG	qRT-PCR
Gata4	CCCAATCTCGATATGTTTGATGAC	ATTACATACAGGCTCACCCTC	qRT-PCR
Gata6	AGGATGTGACTTCGGCAGG	GCATCAGTGATGTCTGCAGT	qRT-PCR
Gooseco id	TGCACCTTCGGGAGGAGAAG	CCGAGGAGGATCGCTTCTGT	qRT-PCR
Hex	CGGACGGTGAACGACTACAC	CGTTGGAGAACCTCACTTGAC	qRT-PCR
Hnf1b	CCCCTCACCATCAGCCAAG	GGTTCTGAGATTGCTGGGGATT	qRT-PCR
Hnf4a	GGTCAAGCTACGAGGACAGC	ATGTACTTGGCCCACTCGAC	qRT-PCR
IAP 1st	TTGATAGTTGTGTTTTAAGTGGTAAAA	AAAACACCACAAACCAAAATCTTC	Bisulphite
IAP 2st	TTGTGTTTTAAGTGGTAAATAAATAAG	CAAAAAAAACACACAAACCAAA	Bisulphite
Krt7	ACCCTCAACAACAAATTCGCGTCC	TGCTCTTGGCTGACTTCTGTTCCT	qRT-PCR
Mixl1	ACCACCAGGCCTGACAACCT	TGGGTGCACACCATACCACA	qRT-PCR
MuErvL	GGCGCATCTGCGCACCTAAA	TAGGGTTAGACACCGGGGTT	qRT-PCR
Nanog	GAGTGTGGGTCTTCCTGGTC	GAGGCAGGTCTTCAGAGGAA	qRT-PCR
Nodal	CAGAATTGCGCCGGGATT	CCGGGATTACCAGAATTGCG	qRT-PCR
Nr0b1	TACCATCTCCTCCCTAGGG	CACCAAATCCCGGCCGGTTT	qRT-PCR
Oct4	CCAGGCAGGAGCACGAGTGG	CCACGTCGGCCTGGGTGTAC	qRT-PCR
Otx2	GGGCCACCAGGACTTACGGT	TATACCGCCGGGCCACCA	qRT-PCR
Pdgfra	GCAGCCCACACCGGATGGTA	TCCGGATCTGTGGTGCGGCA	qRT-PCR
Pecam1	CAAAGTGGAATCAAACCGTATCT	CTACAGGTGTGCCCGAG	qRT-PCR
Rex1	GCTCCTGCACACAGAAGAAA	GTCTTAGCTGCTTCCTTCTTGA	qRT-PCR
Sox17	CACAACGCAGAGCTAAGCAA	TTGTAGTTGGGGTGGTCCTG	qRT-PCR
Sox2	CTGTTTTTTCATCCCAATTGCA	CGGAGATCTGGCGGAGAATA	qRT-PCR
Sox7	CTTCAGGGGACAAGAGTTCG	GCTTGCCTTGTTTCTTCCTG	qRT-PCR
T-Bra	GTCAGACCAAGATCGCTTCT	GATCGCTTCTGTCAGACCAA	qRT-PCR
Tet1	GAAGCTGCACCCTGTGACTG	GACAGCAGCCACACTTGGTC	qRT-PCR
Tet2	AAGCTGATGGAAAATGCAAGC	GCTGAAGGTGCCTCTGGAGT	qRT-PCR
Tet3	TCACAGCCTGCATGGACTTC	ACGCAGCGATTGTCTTCCTT	qRT-PCR
Thbd	CTTCAGGGGACAAGAGTTCG	GCTTGCCTTGTTTCTTCCTG	qRT-PCR
Ttr	TTTCTTCCTGGCTTGCCTTG	AGGATGACCA CTGCTGACTG	qRT-PCR

Table of antibodies used for FACS, WB, immunofluorescence (IF) and dot blot.

ANTIBODY	CATALOGUE N°	COMPANY	DILUITION	APPLICATION
PDGFRα-APC	17-1401-81	ebioscience	1µl/10^6 cells	FACS , IF
PECAM1-FITC	11-0311-85	eBioscience	1µl/10^6 cells	FACS, IF
rIgG2a-APC	17-4321-81	ebioscience	1µl/10^6 cells	FACS
rIgG2a-FITC	11-4321-82	eBioscience	1µl/10^6 cells	FACS
Mouse anti-Oct3/4-488	560217	BD	1µl/10^6 cells	FACS
Active Casp3-647	560626	BD	1µl/10^6 cells	FACS
Ki67-647	558615	BD	5µl/10^6 cells	FACS
Gata4-PE	560328	BD	20µl/10^6 cells	FACS
Anti-Mouse Nanog -APC	50-5761-82	ebioscience	5µl/10^6 cells	FACS,WB
mIgG1-PE	550617	BD	1µl/^{106} cells	FACS
B-tubulin	05-661	Millipore	1/1,000	WB
Oct3/4 (N-19)	SC-8628	Santa Cruz B.	1/1,000	WB
Sox2	AB5603	Millipore	1/1,000	WB
Gata4 (C-20)	SC-1237	Santa Cruz B.	1/1,000	WB
Gata6	AF1700	R&D systems	1/1,000, 1/400	WB, IF
Donkey anti-Ms IgG-HRP	SC-2314	Santa Cruz B.	1/3,000	WB
Donkey anti-Rb IgG-HRP	SC-2313	Santa Cruz B.	1/3,000	WB
Donkey anti-Gt IgG-HRP	SC-2056	Santa Cruz B.	1/3,000	WB
Rb anti-5hmC	39791	Active Motif	1/5,000	Dot Blot
Ms anti-5mC	39649	Active Motif	1/500	Dot Blot
Lamininβ2	DSHB -D18	Hybridoma Bank	1/400	IF
Cdx2	Cdx2-88	Biogenex	1/200	IF
Foxa2	AB40874	Abcam	1/100	IF
Mixl1	ABS232	Millipore	1/100	IF
AlexaFluor 488 Donkey anti-Gt	A11055	Invitrogen	1/500	IF
AlexaFluor 555 Donkey anti-Gt	A21432	Invitrogen	1/500	IF
AlexaFluor 488 Donkey anti-Ms	A21202	Invitrogen	1/500	IF
AlexaFluor 555 Donkey anti-Rb	A21206	Invitrogen	1/500	IF

SUPPLEMENTAL REFERENCES

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193-196.

http://hannonlab.cshl.edu/fastx_toolkit/index.html.

http://www-huber.embl.de/users/anders/HTSeq.

http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

M. Morgan, M.L., and S. Anders. (2009). ShortRead: Base classes and methods for high-throughput short-read sequencing data.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A.*, et al.* (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590-604.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17(1).

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. BMC bioinformatics *12*, 480.

Robinson, M.D., and Smyth, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance. Bioinformatics *23*, 2881-2887.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology *28*, 511-515.