

Defining the nature of human pluripotent stem cell progeny

Michaela Patterson^{1,3,*}, David N Chan^{1,3,*}, Iris Ha¹, Dana Case⁴, Yongyan Cui¹, Ben Van Handel^{1,3},
Hanna KA Mikkola^{1,2,3,4}, William E Lowry^{1,2,3,4}

¹Department of Molecular, Cell and Developmental Biology; ²Jonsson Comprehensive Cancer Center; ³Eli and Edythe Broad Center for Stem Cell Research; ⁴Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

While it is clear that human pluripotent stem cells (hPSCs) can differentiate to generate a panoply of various cell types, it is unknown how closely *in vitro* development mirrors that which occurs *in vivo*. To determine whether human embryonic stem cells (hESCs) and human-induced pluripotent stem cells (hiPSCs) make equivalent progeny, and whether either makes cells that are analogous to tissue-derived cells, we performed comprehensive transcriptome profiling of purified PSC derivatives and their tissue-derived counterparts. Expression profiling demonstrated that hESCs and hiPSCs make nearly identical progeny for the neural, hepatic, and mesenchymal lineages, and an absence of re-expression from exogenous reprogramming factors in hiPSC progeny. However, when compared to a tissue-derived counterpart, the progeny of both hESCs and hiPSCs maintained expression of a subset of genes normally associated with early mammalian development, regardless of the type of cell generated. While pluripotent genes (*OCT4*, *SOX2*, *REX1*, and *NANOG*) appeared to be silenced immediately upon differentiation from hPSCs, genes normally unique to early embryos (*LIN28A*, *LIN28B*, *DPPA4*, and others) were not fully silenced in hPSC derivatives. These data and evidence from expression patterns in early human fetal tissue (3-16 weeks of development) suggest that the differentiated progeny of hPSCs are reflective of very early human development (< 6 weeks). These findings provide support for the idea that hPSCs can serve as useful *in vitro* models of early human development, but also raise important issues for disease modeling and the clinical application of hPSC derivatives.

Keywords: hESC; hiPSC; differentiation; human development; human fetal tissue; pluripotent
Cell Research (2012) 22:178-193. doi:10.1038/cr.2011.133; published online 16 August 2011

Introduction

hiPSCs are similar to hESCs in that they share the same pluripotency markers, display self-renewal, and are capable of differentiation into all three germ layers. The rapid development of hiPSC technology has delivered new hope for personalized medicine. In addition, hiPSCs may present a better tool for modeling disease as they can be derived from patients with various genetic diseases at virtually any age [1].

We and others have demonstrated that undifferentiated hiPSCs at early passages display a somewhat unique gene expression pattern from hESCs [2-4]. However, the

functional relevance of any gene expression differences between hiPSCs and hESCs remains unclear. Recent evidence suggests that these differences may be explained by the fact that hiPSCs are more variant than hESCs at the epigenetic and transcriptional level [5-7]. It is also possible that the differences between hiPSCs and hESCs persist upon differentiation and that these differences influence the derivation of specified progeny, but this has not been formally tested. Recent work in murine and human iPSCs has shown that a residual epigenetic memory of the starting cell type is retained after reprogramming, and this memory appears to influence the efficiency of differentiation [8-11]. Furthermore, epigenetic differences between mouse iPSC lines diminish, as the cells are passaged [8], consistent with what was shown for gene expression between hiPSCs and hESCs [2, 12]. So far, differentiation efficiency appears to be as variable amongst hESC lines as it is for hiPSC lines [5, 13-15].

Despite varying efficiencies, all hPSC lines appear to have some potential to generate an array of functional

*These two authors contributed equally to this work.

Correspondence: William E Lowry

Tel: +1 310-794-5175; Fax: +1 310-794-9323

E-mail: blowry@ucla.edu

Received 19 April 2011; revised 1 June 2011; accepted 6 July 2011; published online 16 August 2011

cell types, including motor neurons, cardiomyocytes, dopaminergic neurons, hematopoietic cells, hepatocytes, etc. [16-21]. While it is clear that hiPSCs are capable of generating similar cell types as hESCs as determined by expression of a select number of markers, it remains to be seen if, on a global level, these derivatives are truly analogous. Furthermore, for hESCs or hiPSCs to be employed in therapy or as models of human development, it is also important to determine how well their differentiation mirrors natural development. Because both hiPSCs and hESCs undergo development *in vitro* instead of *in vivo*, it is not clear if current differentiation protocols accurately recapitulate the development that occurs in the human embryo. Nor is it clear whether cell types can be generated from hPSCs that mimic mature adult phenotypes.

To address these issues, we generated and purified derivatives of all three germ layers from hESCs and hiPSCs using established protocols. We compared the global gene expression pattern between the progeny of hESCs and hiPSCs, and primary tissue-derived equivalents from various stages of development. Our results show that the gene expression differences found between hiPSCs and hESCs in the undifferentiated state mostly dissipated after differentiation and derivatives from hiPSCs and hESCs were very similar to each other transcriptionally. On the other hand, a significant number of genes were different between PSC derivatives and their respective natural counterpart, regardless of the cell type generated. While *OCT4*, *SOX2*, *REX1*, and *NANOG* were effectively silenced, an alternative set of genes normally associated with the pluripotent state including *LIN28A*, *LIN28B*, and *DPPA4* remained expressed in pluripotent cell-derived progeny. We hypothesize and provide significant evidence that these cells derived from hPSCs are representative of cells found during human development prior to 6 weeks of development. Our data are consistent with the notion that ESC-derived cells represent early stages of development [22-25], but we demonstrate here that this also applies to hiPSC-derived cells. Furthermore, we attempt to place the PSC derivatives into a more precise developmental time frame than that has been established previously. These findings underscore the need to consider the maturity of cells produced from hPSCs for disease modeling or regenerative medicine.

Results

hESC and hiPSC lines are capable of generating derivatives representing all three embryonic germ layers

Using previously established protocols, we differentiated hESCs and hiPSCs into derivatives representing

all three embryonic germ layers: neural progenitor cells (NPCs) (ectoderm) [18], hepatocytes (endoderm) [20], and fibroblasts (FBs) (mesoderm). In our studies, the efficiency of derivation of various cell types from PSCs (including hESCs and FB-derived hiPSCs) was highly variable amongst various lines and across multiple experiments (data not shown and Karumbayaram *et al.* [18]). However, it was clear that all the PSC lines used here were able to generate a reasonable number of cells of the indicated type for analysis of gene expression and functional capacity.

Generation of ectoderm

When hESCs and hiPSCs were directed to generate NPCs and were isolated based on rosette morphology (p1), well-established neural stem cell markers (MUSASHI, PROMININ, PAX6, SOX2, and NESTIN) were induced as measured by RT-PCR (Supplementary information, Figure S1A). Immunocytochemistry confirmed the expression of these and other NPC markers at the protein level (Figure 1A and Supplementary information, Figure S1A'). By immunostaining, all the NPC markers labeled at least 80% of cells, demonstrating that the culture represented a homogenous pool of NPCs. While hESC and hiPSC-derived NPCs appeared to express these genes at a similar level to one another, there was variation in the level of protein expression on a per-cell basis observed between PSC-NPCs and those isolated from 16-week-old fetal brain. The transcription factor PAX6 was expressed at a lower level in PSC-NPCs when compared to 16-week-old fetal brain-derived NPCs (FNPC-16 Br) on a per-cell basis, while NESTIN expression was higher (Figure 1A). Because retinoic acid and smoothed agonist were used in the neural specification, it is possible that we induced a more posterior/ventral fate than NPCs isolated from fetal brain. To explore this possibility, the PCS-NPCs were also compared with NPCs isolated from 15.5-week-old fetal spinal cord (FNPC-15.5SC) and expanded under the same conditions. In fact, NESTIN and PAX6 were expressed in FNPC-15.5SCs at a similar level as in our PCS-NPCs (Figure 1A and Supplementary information, Figure S1A'). By immunostaining and morphology, both PSC-NPCs and 16-week-tissue-derived NPCs had the capacity to generate both Tuj1+ neurons and GFAP+ glia (Figure 1A'); however, the p1 PSC-NPCs mostly generated neurons (Tuj1+), while the 16-week-old-tissue-derived NPCs mostly produced glia (GFAP+) (Figure 1A''). Because neurogenesis precedes gliogenesis during *in vivo* development [26], these data suggest that PSC-NPCs may represent earlier developmental time points than the NPCs derived from 16-week-old fetal tissue.

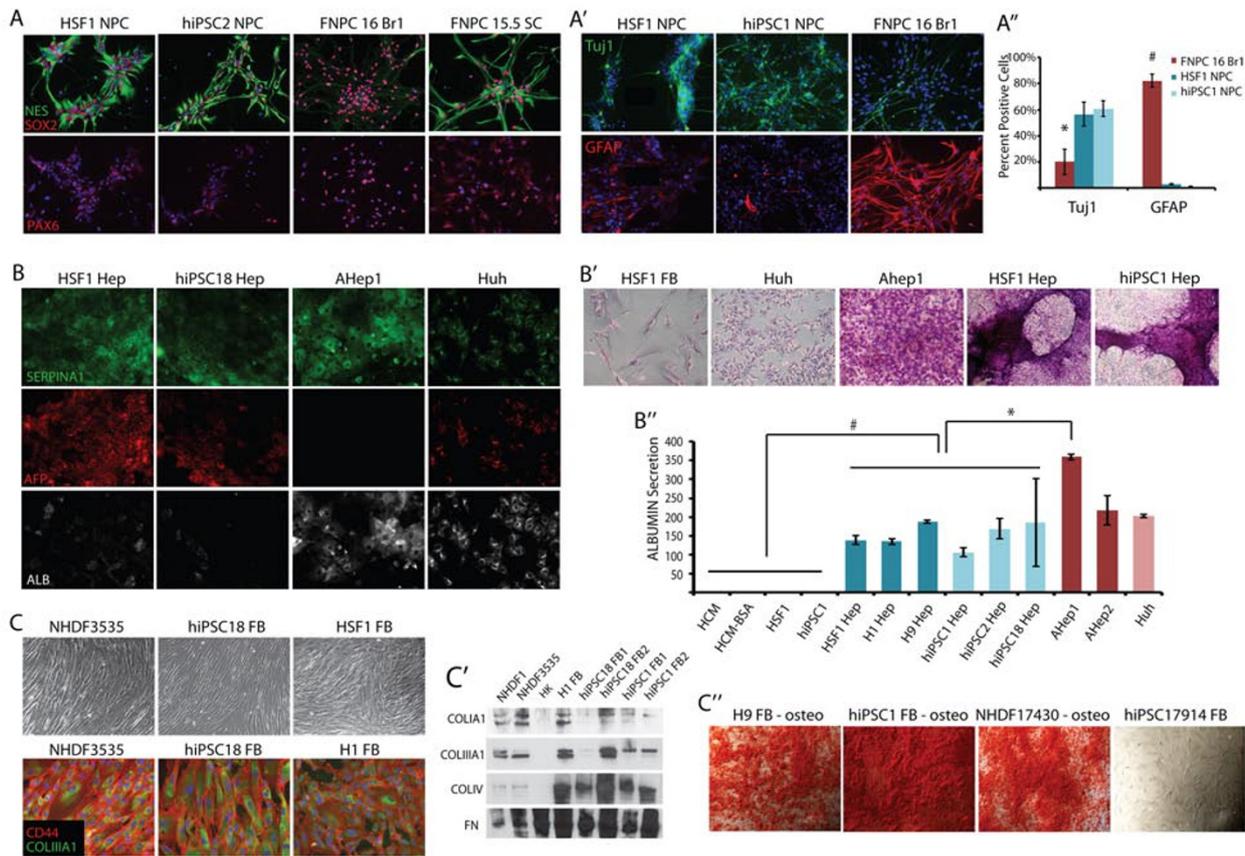


Figure 1 hESC and hiPSC lines make cell types representing all three germ layers. hESC and hiPSC lines were directed to differentiate into either NPCs (**A**), hepatocytes (**B**), or fibroblasts (**C**). (**A**) Immunofluorescence staining for SOX2 (red, top panel), NESTIN (green, top panel), and DNA (blue) and PAX6 (red, bottom panel). (**A'**) Immunofluorescence staining demonstrating that NPCs derived from hESCs, hiPSC, or natural sources could be differentiated into Tuj1+ neurons (green) and GFAP+ glia (red). (**A''**) Quantification of the percent of cells undergoing neuronal (Tuj1+) or glial (GFAP+) differentiation. Error bars represent standard error over 5-8 fields of view. * $P < 0.05$; # $P < 1.0E-06$. (**B**) Immunofluorescence staining of SERPINA1 (green), AFP (red), ALBUMIN (white), and DNA (blue). (**B'**) Periodic acid-schiff assay stain demonstrating glycogen storage in natural- and pluripotent-derived hepatocytes. (**B''**) ELISA measuring albumin secretion on confluent plates. Error bars represent standard error over two replicates. * $P < 0.01$; # $P < 0.05$. (**C**) Top, phase-contrast images of fusiform morphology displayed by pluripotent- and naturally derived fibroblasts. (**C**) Bottom, immunofluorescence staining of CD44 (red), COL1A1 (green), and DNA (blue). (**C'**) Western blot for secreted collagen proteins (COLIA1, COL1A1A1, and COLIV) and FIBRONECTIN (FN). HK, human keratinocyte. (**C''**) Alizarin Red stain following further differentiation of pluripotent cell- and tissue-derived fibroblasts down the osteogenic lineage.

Generation of endoderm

For hepatocyte derivation, hESCs and hiPSCs were directed to undergo definitive endoderm formation, hepatic specification, hepatoblast expansion, and finally hepatic maturation as described previously [16, 20]. After 3 days of differentiation, definitive endoderm markers, *HNF3B* and *SOX17*, were induced, as demonstrated by RT-PCR (Supplementary information, Figure S1B). Following hepatic induction and expansion, the *SOX17* mRNA level declined, while various hepatic markers (*AFP*, *ALB*, *SERPINA1*, *CYP3A4*, and *CYP3A7*) were induced over

time, as found in mature adult tissue-derived hepatocytes or a hepatocarcinoma cell line (Figure 1B and Supplementary information, Figure S1B). By morphology and immunostaining, the hepatocytes produced from PSCs were more similar to hepatoblasts, or immature hepatocytes that populate the developing early fetal liver (Figure 1B and Supplementary information, Figure S1B). These hepatic derivatives expressed a higher level of fetal hepatic genes such as *AFP* and *CYP3A7*, and a lower level of the more mature equivalents *ALBUMIN* and *CYP3A4*, when compared to counterparts made from adult liver

(Figure 1B and Supplementary information, Figure S1B). Again, this suggested that PSC-Heps could represent a younger developmental stage than those of adult liver. Regardless, the hepatocytes generated from PSCs were able to both store glycogen (Figure 1B') and secrete albumin (Figure 1B'').

Generation of mesoderm

For FB generation, embryoid bodies (EBs) were first generated from hESCs and hiPSCs. The EBs were then plated in adherent conditions and grown in FB culturing media. Following several passages, these cultures began to display a homogenous fusiform morphology typical

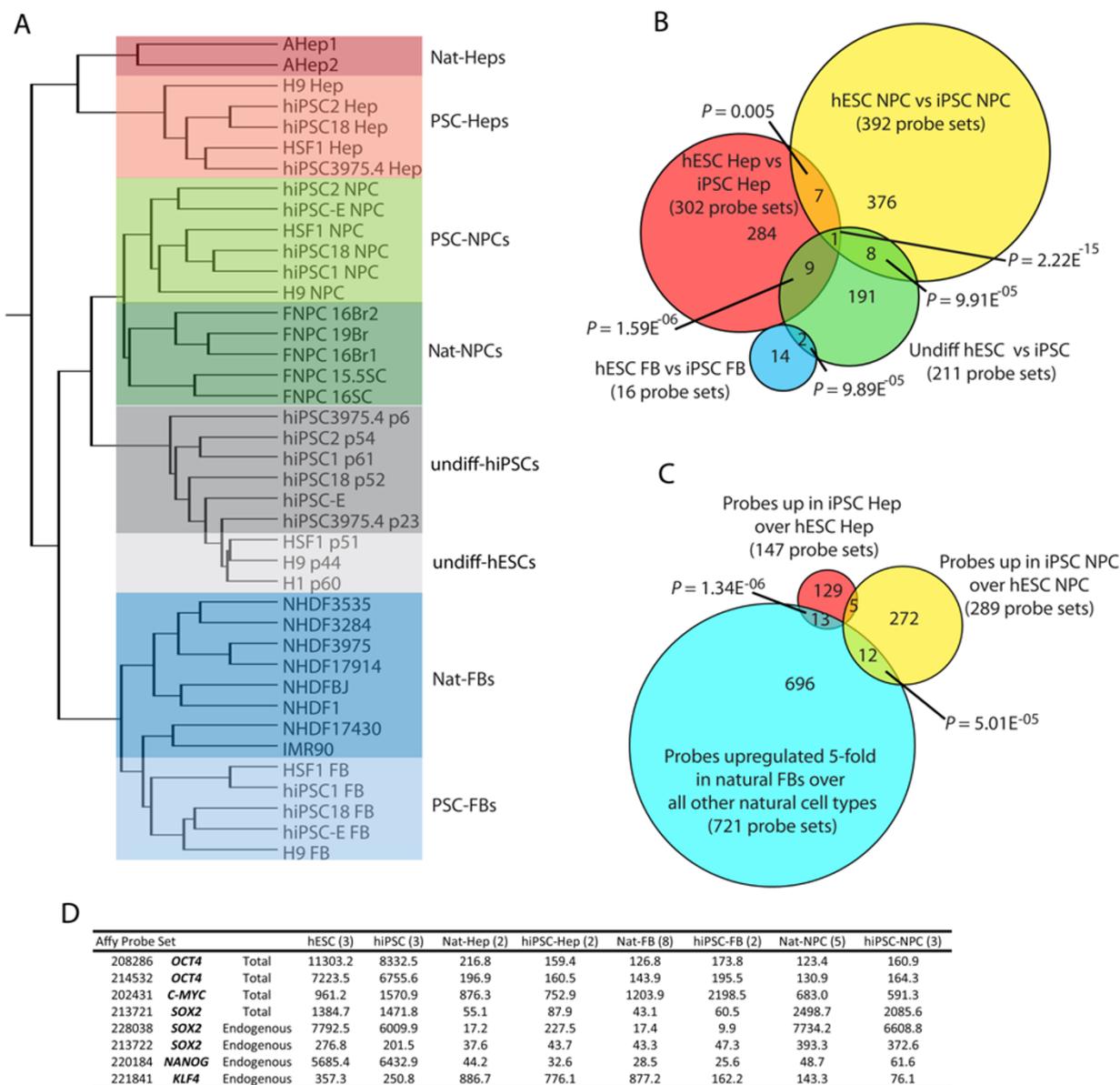


Figure 2 Global gene expression analysis. **(A)** Hierarchical clustering analysis of global gene expression in undifferentiated hESCs, hiPSC, and their progeny compared to naturally derived cells. **(B)** Venn diagram summarizing the probe sets that were differentially expressed (t -test $P < 0.01$; fold change ≥ 1.54) between the progeny of hiPSCs versus the progeny of hESCs for each germ layer and the undifferentiated. **(C)** Venn diagram overlapping fibroblast signature probe sets (t -test between natural-FB and all other natural cell types; upregulated in FBs ≥ 5.0) with probe sets upregulated in iPSC progeny over ESC progeny for the NPC and Hep lineages. P -Values from **B** and **C** were measured by hypergeometric distribution or simulation as in [2]. **(D)** Normalized values from microarray probe sets for the reprogramming factors used to make the hiPSCs used in this study.

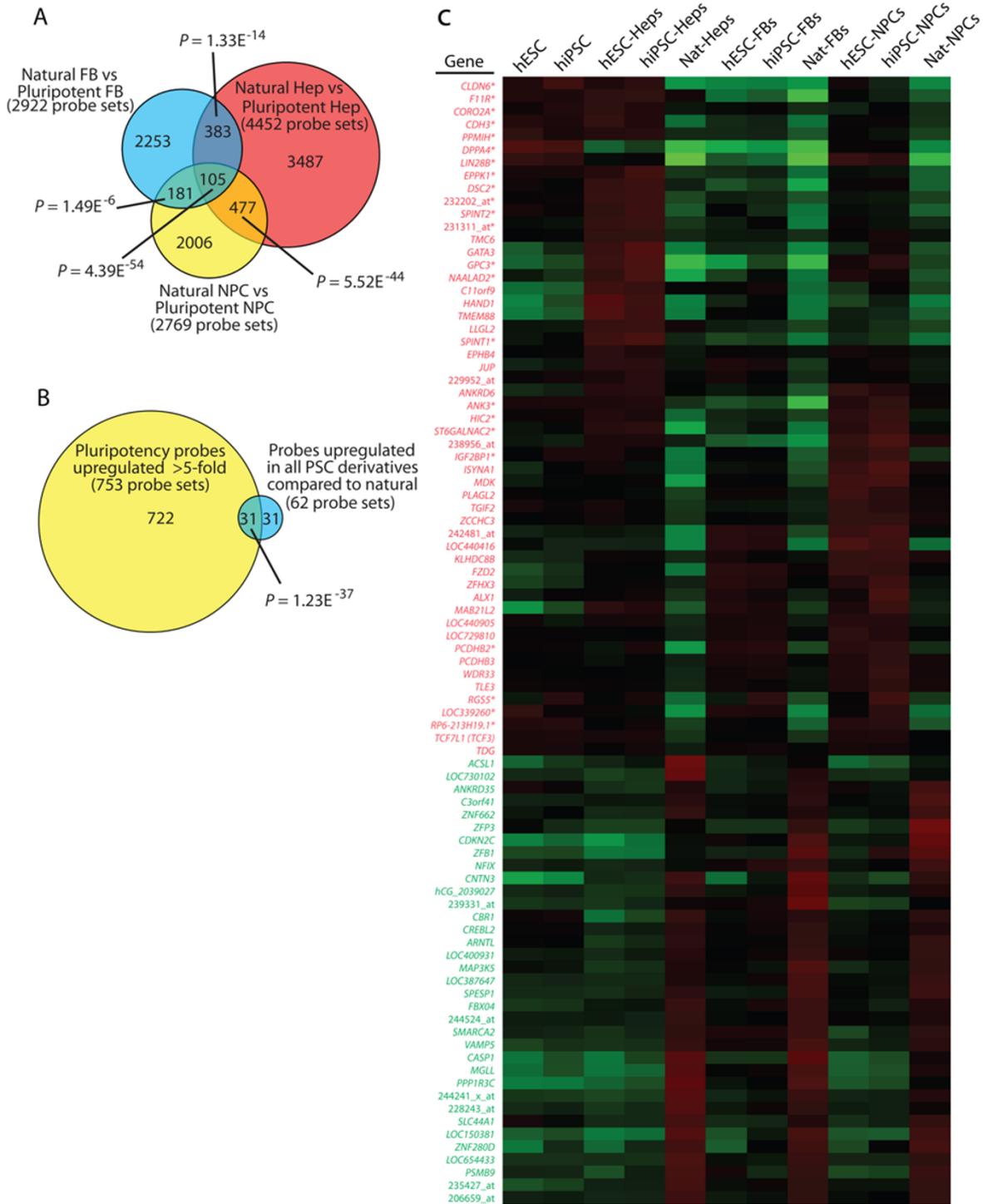


Figure 3 Expression profiling identifies a conserved list of probe sets differentially expressed between pluripotent derivatives and their natural counterparts. **(A)** A *t*-test ($P < 0.01$) was performed to identify probe sets differentially expressed between PSC derivatives (FB, blue; Hep, red; NPC, yellow) and their respective natural counterparts (fold change ≥ 1.54). Venn diagram reveals the overlap of these differences across the different progeny. **(B)** Overlap of the 62 probe sets specifically upregulated in 3A with probe sets that demonstrate a significant difference between pluripotent cells and naturally derived somatic cells (fold upregulation ≥ 5). *P*-Values from **A** and **B** measured by hypergeometric distribution or simulation as in [2]. **(C)** A heat map was generated for the 88 unique genes or ESTs represented by the 105 probe sets shown in **A** that are differentially expressed between PSC derivatives and tissue derived cells. Note that signal shown represents value divided by the average of all samples and genes in red were consistently found upregulated in PSC progeny versus tissue-derived cells, while those in green were always downregulated. *Indicates genes expressed highly in the pluripotent stem cells (identified in **B**).

of FBs (Figure 1C). For comparison, FBs were derived from human dermis or lung at various developmental time points and were grown *in vitro* under the same conditions. When characterized at the RNA level, FB markers were expressed in PSC-FBs at a level comparable to FBs derived from the dermis of skin (Supplementary information, Figure S1C). Furthermore, using immunocytochemistry, two fibroblast markers, CD44 and COL1A1, were expressed in PSC-FBs at levels comparable to that observed in a neonatal dermal FB line (Figure 1C). Functionally, the PSC-FBs secreted a profile of collagens similar to that secreted by dermal FBs, while human keratinocytes did not (Figure 1C'). In addition, the PSC-FBs, like their natural counterparts, were not necessarily terminally differentiated, as they still retained the ability to undergo osteogenic specification (Figure 1C'').

Global characterization of PSC derivatives by gene expression profiling

To more precisely determine the identity of the derivatives of PSCs, each of the indicated cell types was purified and profiled for gene expression. We analyzed the transcriptional profiles of PSC derivatives as well as natural counterparts, and performed unsupervised hierarchical clustering (Figure 2A). From this clustering analysis we made several important observations.

First, most of the gene expression differences observed between hiPSCs and hESCs in the undifferentiated state were not found when hESC and hiPSC were differentiated, evident by the fact that hESC- and hiPSC-derived progeny did not segregate from one another. Second, while a small number of significant differences was detected between hESC progeny and that of hiPSCs for each lineage (roughly 300), these differences were not detected when a more stringent statistical measure (false discovery rate (FDR)) was applied (Figure 2B and data not shown). Furthermore, very few of these differences overlapped with the list of genes differentially expressed between undifferentiated hESCs and hiPSCs (Figure 2B), and none of them could be categorized with gene ontological (GO) analysis. Whether these differences serve to functionally distinguish hESC from hiPSC derivatives will require extensive investigation of differentiated progeny from hiPSCs, reflecting a variety of different somatic cell types.

Recent evidence suggests that murine and human iPSCs retain an epigenetic signature from their cell of origin that influences their differentiation potential [8-11]. All of the hiPSCs used in our analysis were derived from FBs, making similar types of analysis in our model system impossible. We did, however, look for residual expression of FB-specific genes in hiPSC derivatives ver-

sus hESC derivatives and found a handful of genes that might represent residual gene expression from the target cell of reprogramming (Figure 2C). We also ruled out that any gene expression difference was due to expression of integrated reprogramming factors in the hiPSCs because the reprogramming factors were not significantly expressed in these progeny, except for *SOX2*, as would be expected for NPCs (Figure 2D).

The most striking observation made from our transcriptome analysis was that, while PSC derivatives representing the three germ layers clustered with their respective natural counterparts, within each germ layer cluster, the PSC progeny were always distinguished from their naturally derived counterparts (Figure 2A). This finding suggested that, with the methods employed here and commonly applied elsewhere, PSC derivatives are similar but not identical to tissue-derived counterparts. This led to further investigation of the transcriptional differences between hPSC derivatives and their tissue-derived counterparts, and whether these differences are reflective of differences in developmental maturity or arise due to development performed *in vitro*.

PSC derivatives and tissue-derived counterparts are distinguished by their gene expression

The genes differentially expressed between PSC derivatives and their tissue-derived counterparts were compared using a *t*-test ($P < 0.01$) and requiring at least a 1.54-fold expression difference. Out of 36 749 probe sets that were expressed in at least 10% of samples, 2 922 were differentially expressed between PSC-FBs and dermal/lung FBs; 4 452 were differentially expressed between PSC-Heps and adult hepatocytes; and 2 769 were differentially expressed between PSC-NPCs and 16-week-old fetal NPCs. GO analysis of the genes differentially expressed in each germ layer representative yielded many categories for hepatocytes, but few for NPCs and FBs (Supplementary information, Figure S2D). Surprisingly, when superimposing these pools of differentially expressed probe sets and taking direction of differential expression into account, 105 were found to be differentially expressed between all PSC derivatives and their natural counterparts (Figure 3A), suggesting that all types of PSC derivatives share common differences with tissue-derived cells. No GO terms were conserved across these 105 probe sets (Supplementary information, Figure S2D). A more stringent analysis employing FDR correction produced a smaller list of probe sets, but the differences still overlapped significantly across the three germ layers and were entirely inclusive with the non-FDR corrected analysis (Supplementary information, Figure S2A).

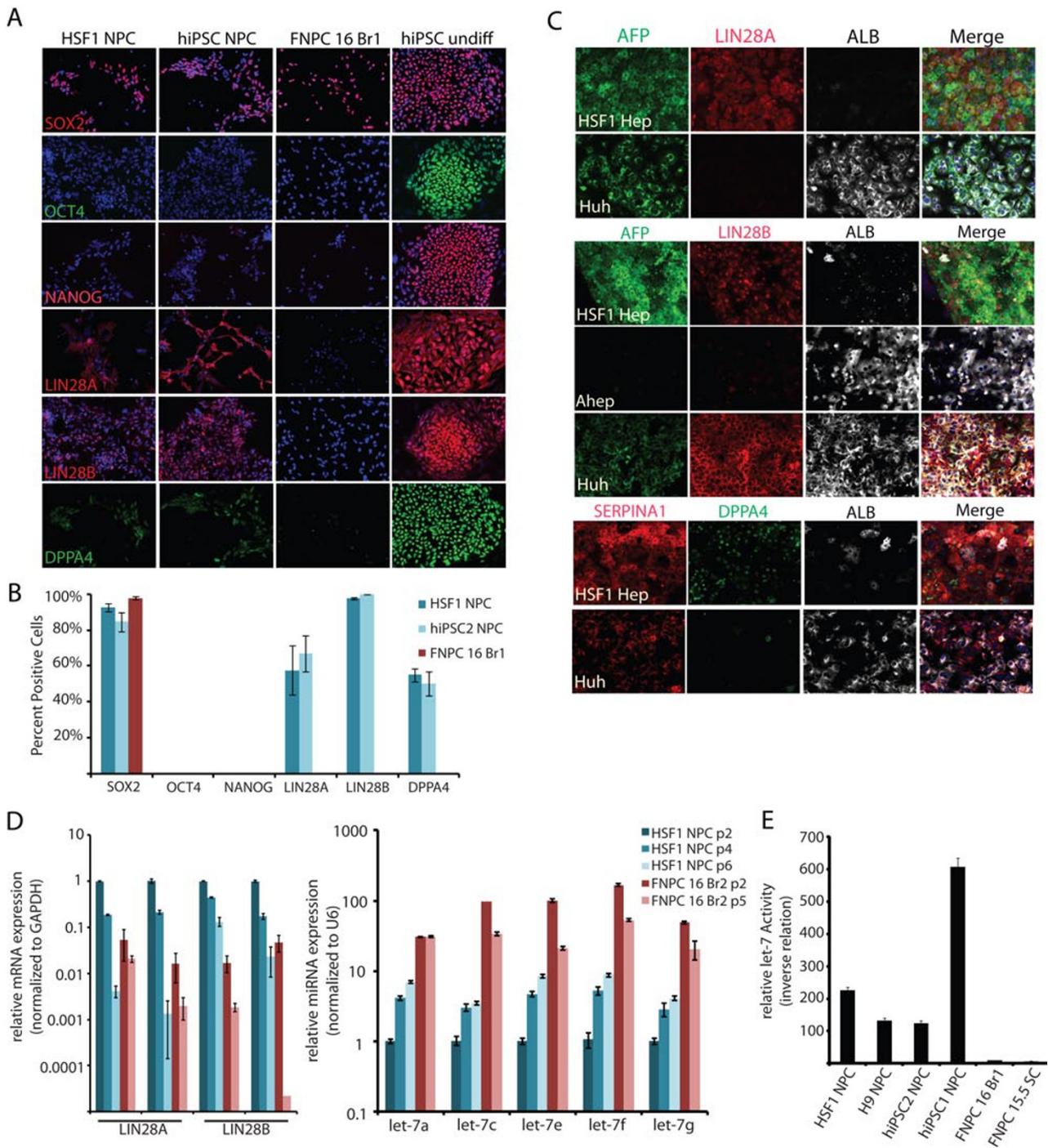


Figure 4 Expression and activity of LIN28 and DPPA4 in PSC derivatives. **(A)** NPCs made from PSCs and brain were stained with pluripotency markers SOX2, OCT4, NANOG, LIN28A, LIN28B, and DPPA4. Undifferentiated hiPSCs were stained as a positive control for the pluripotency markers. **(B)** Quantification of the percent of FNPCs and PSC-NPCs expressing the indicated pluripotency markers. **(C)** HSF1-derived hepatocytes and control cells were immunostained with antibodies recognizing ALBUMIN, AFP, or SERPINA1 to highlight both immature and mature cells and either LIN28A, LIN28B, and DPPA4 to demonstrate that these pluripotency factors are not silenced immediately upon differentiation. Hepatocytes taken from adult human liver did not express any of these pluripotency genes, while Huh, a hepatocarcinoma cell line expressed LIN28B. **(D)** Real-time PCR for *LIN28A* and *LIN28B* mRNA (left) and *let-7* miRNA family members (right). mRNA expression was normalized to GAPDH, while miRNA expression was normalized to U6. Error bars represent standard error over three or four replicates. **(E)** To determine the relative let-7 activity in the indicated cell types, each was transfected with let-7 reporter and constitutive reporter as a transfection control. Dual luciferase assays were performed 48 h after transfection in triplicate. Assay shown was representative of three independent experiments.

From this pool of 105 differentially regulated probe sets, 62 were upregulated in all PSC derivatives versus their tissue-derived counterparts. Notably, 31 of these 62 probe sets overlap with probe sets that are highly upregulated in undifferentiated PSCs versus specified somatic cells (Figure 3B). This indicates that the PSC progeny continued to express a significant subset of genes associated with either pluripotency or early embryonic development. Many of these same genes were also expressed at a higher level in PSC derivatives versus tissue-derived cells generated independently by other groups, suggesting that these observations were not specific to our methods or the particular cell lines used here (Supplementary information, Figure S2C). *OCT4*, *SOX2*, and *NANOG* were not amongst the 31 probe sets related to pluripotency that remained high in PSC derivatives, demonstrating that these genes were silenced upon differentiation, as has been shown extensively. Instead, most of the 31 probe sets appeared to be not only expressed in PSCs but also play roles in early embryonic development as judged by functional data from lower organisms [27-29]. The expression pattern of all genes specifically expressed in PSCs can be found in Supplementary information, Table S1.

The expression pattern across cell types for the 88 unique genes and unannotated probe sets differentially expressed between PSC derivatives and tissue-derived counterparts (represented by 105 probe sets) is represented in Figure 3C as a heat map. Of the 53 genes that were higher in PSC derivatives (in red), 22 were also strongly expressed in undifferentiated PSCs relative to somatic cells (indicated with asterisk). This list included *LIN28B*, *DPPA4*, and *TCF7L1* (*TCF3*), all of which are known to play a role in ESCs and in very early mammalian development [28, 30-33]. Furthermore, 35 genes were downregulated in PSC derivatives compared to tissue-derived cells (in green), perhaps reflecting a state of incomplete specification, regardless of the cell type generated.

LIN28 was first discovered as a regulator of developmental timing in *C. elegans* [29, 34]. *LIN28A* and *LIN28B* are highly expressed in undifferentiated hPSCs, but are thought to be silenced, as tissues are specified and mature. *LIN28A* has also been employed as a reprogramming factor in the generation of hiPSCs [35], suggesting that it can play a functional role in maintaining or inducing immature cell fate. *LIN28B* was expressed in all hESC- and hiPSC-derivatives, whereas *LIN28A* was found at a high level in PSC-NPCs and PSC-Heps, but not in PSC-FBs. *LIN28A*, *LIN28B*, and *DPPA4* were also frequently expressed at a high level in PSC derivatives generated by other groups (Supplementary information, Figure S2C), indicating that expression of these genes

is not confined to just the hESC and hiPSC derivatives produced in our lab, but is perhaps an attribute of PSC derivatives in general.

To determine whether the changes detected in the RNA profiling data were just due to residual-stabilized RNA from the pluripotent state, the expression of *LIN28A*, *LIN28B*, and *DPPA4* in PSC derivatives was also examined at the protein level. In fact, *DPPA4*, *LIN28A*, and *LIN28B* were all expressed at the protein level in PSC-NPCs (Figure 4A and 4B) and PSC-Heps (Figure 4C), but not in their tissue-derived counterparts. On the other hand, none of the other classic pluripotency factors (*OCT4*, *NANOG*, and *REX1/ZFP42*) were expressed in any of the PSC derivatives at the RNA or protein levels (Figure 4A-4C, and data not shown).

The LIN28/let-7 circuit in PSC derivatives

LIN28 has been shown to act as an RNA-binding protein that regulates miRNA maturation, particularly the *let-7* family [27, 29-31, 36-40]. If *LIN28* activity is higher in PSC derivatives, these cells would be expected to have low levels of mature *let-7*. To determine if the increased *LIN28* expression in PSC-NPCs correlated with low levels of mature *let-7*, RT-PCR was used to probe the relative levels of mature *let-7* miRNA family members. As expected, PSC-NPCs were found to have very low relative levels of mature *let-7* family members compared to tissue-derived NPCs (Figure 4D).

To determine if the low levels of *let-7* expression in PSC-NPCs also correlated with low activity of these miRNAs, cells were transfected with a reporter that drives constitutive expression of the Renilla luciferase gene with *let-7* seed sequences added to its 3' UTR [41]. Therefore, higher luciferase activity is a result of decreased *let-7* miRNA activity in the cells. NPCs generated from 16-week-old fetal brain or spinal cord showed very little reporter activity, indicating high *let-7* activity, whereas PSC-NPCs displayed high reporter activity, indicating the opposite (Figure 4E). Considering the *LIN28* expression pattern, and the *let-7* activity assay, PSC derivatives not only had high *LIN28* expression, but also high *LIN28* activity, which in turn led to low *let-7* activity. While the *LIN28/let-7* pathway has been implicated in development of many species including mouse [27, 30], this appears to be the first demonstration that human PSC derivatives have high *LIN28* expression and low *let-7* activity. These data further suggest PSC derivatives are developmentally immature compared to their natural counterparts and that caution is warranted in clinical application of these cells, as many human cancers are defined by high *LIN28* and low *let-7* expression [27, 39].

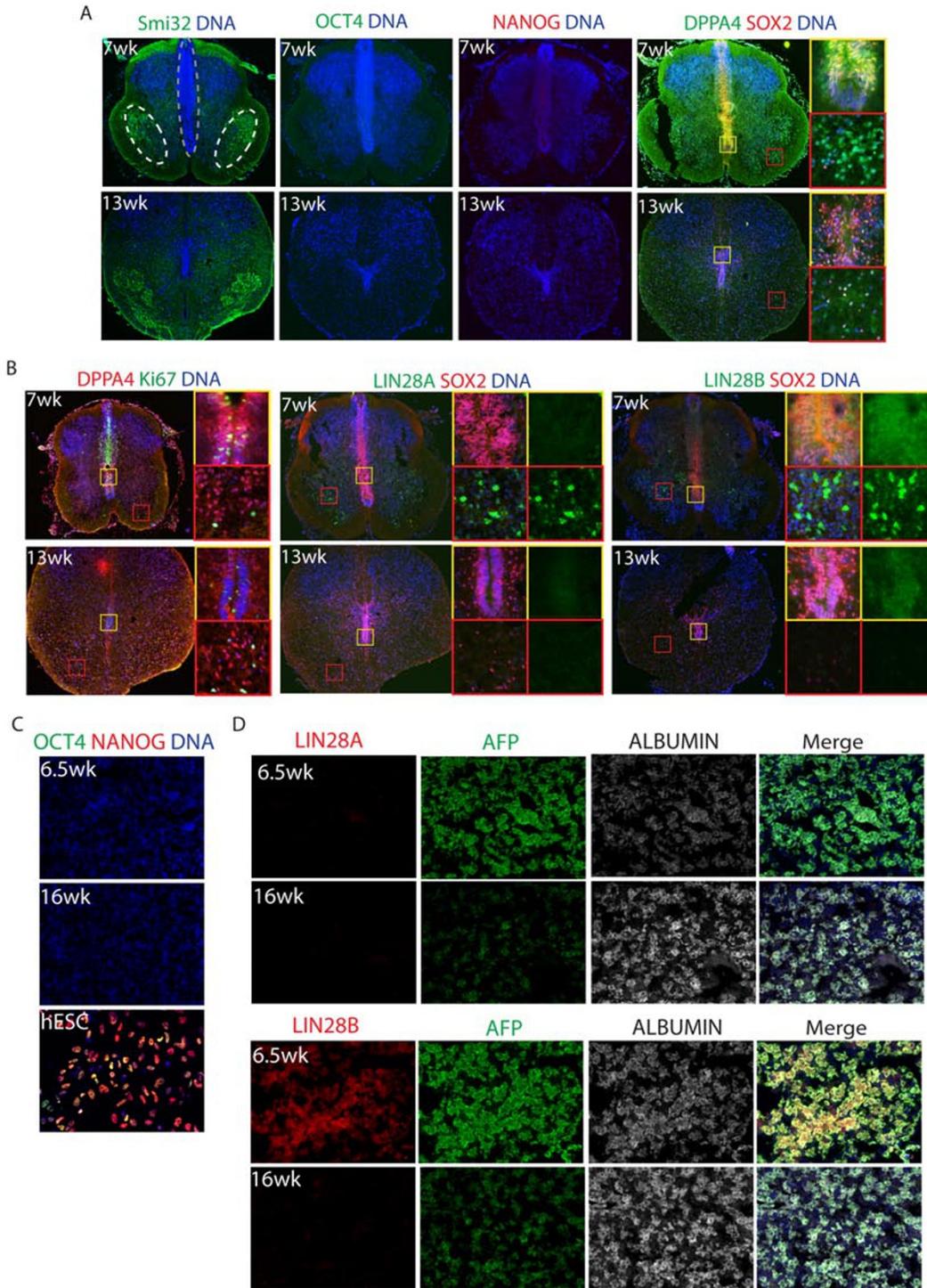


Figure 5 Differentially expressed genes, *DPPA4*, *LIN28A*, and *LIN28B* are found in early fetal tissues. **(A)** Spinal cord tissues (7- and 13-week-old) were fixed, sectioned, and stained with the indicated markers. Smi32 was used to highlight the motor neuron pool (white circles). SOX2 labels the NPCs found along the midline (grey outline). **(B)** *DPPA4* co-localized with proliferation marker Ki67 along the midline (yellow inset) at 7 weeks, but not in the dispersed lateral stain (red inset). At 13 weeks, *DPPA4* was more dispersed and the number of Ki67-positive cells was decreased. *LIN28A* had a cytosolic staining pattern and was located in cells outside of the midline (red and yellow insets) at 7 weeks, but absent at 13 weeks. *LIN28B* had a nuclear staining pattern at 7 weeks. It was often co-localized with SOX2 along the midline at 7 weeks and was weaker at 13 weeks (yellow inset). **(C)** Fetal liver (6.5- and 16-weeks-old) were stained with the indicated pluripotency markers, showing the complete silencing of pluripotency genes. hESCs were used as a positive control. **(D)** Fetal liver (6.5- and 16-week-old) were stained with *LIN28A*, *LIN28B*, and the indicated liver markers.

Expression of early embryonic genes in PSC derivatives suggests that they represent early stages of human development

To determine whether expression of the *LIN28* genes and *DPPA4* in PSC derivatives was indicative of very early fetal development or simply an *in vitro* phenomenon, fetal tissues were also probed for expression of these genes. In the human fetal spinal cord, regional identity was established with *Smi32*, a marker of motor neurons, which was localized to bilateral pools on each side of the ventral cord (Figure 5A). The neural progenitor pool was localized at the midline and expressed *SOX2* and *Ki67* (Figure 5A and 5B). *DPPA4* was also strongly expressed in the midline of the spinal cord along with *SOX2* at 7 weeks of development, but was weaker by 13 weeks (Figure 5B). *LIN28A* was expressed in scattered cells in 7-week-old spinal cord, but was lost by 13 weeks. *LIN28B*, on the other hand, was strongly detected in 7-week-old human spinal cord cells outside of the midline and weakly expressed in the midline progenitor cells along with *SOX2*. Expression of *LIN28B* was significantly reduced, but not absent, in the spinal cord by 13 weeks of development (Figure 5B). The fact that midline progenitor cells of the human spinal cord did express *DPPA4* and *LIN28B* at 7 weeks could further suggest that PSC-NPCs are more similar to an earlier stage of fetal development, or that multiple pools of diverse progenitors are present at this time point.

In the fetal liver, *LIN28B*, but not *LIN28A*, was detectable at 6.5 weeks (Figure 5D). Neither *LIN28* protein was detectable in fetal liver by 16 weeks (Figure 5D). As expected, neither *OCT4* nor *NANOG* was expressed in the spinal cord or fetal liver at any time point analyzed, consistent with the notion that these pluripotent genes are silenced very early in human development (Figure 5A and 5C). These data highlight the possibility that PSC derivatives differed from the tissue-derived cells as shown in Figures 2 and 3, because the cells were taken from tissues that were of at least 16 weeks of development. The presence of both *LIN28* proteins and *DPPA4* in 7-week-old spinal cord and liver, the fact that PSC-NPCs tended to be neurogenic as opposed to gliogenic, and that PSC-Heps express *AFP* instead of albumin, are all consistent with the notion that pluripotent derivatives are similar to cells found at 7 weeks of development or earlier.

Are PSC-NPCs distinguished from fetal NPCs by time in vitro or time in vivo?

Following the observation that early embryonic genes were expressed in younger fetal samples, but significantly reduced at later stages, we were prompted to ask two additional questions: (1) could additional time in culture

bring the PSC derivatives closer to a natural counterpart; and (2) would younger fetal samples more closely resemble our PSC derivatives on a transcriptome level? To answer these questions, we performed a series of additional gene expression profiling experiments on the NPCs.

When PSC-NPCs were cultured for another month (each passage representing ~5-7 days in culture), both *LIN28A* and *LIN28B* mRNA were reduced as measured by RT-PCR analysis (Figure 4D). Furthermore, let-7 family members were upregulated with extended culture (Figure 4D). The reduction of *LIN28A* and *LIN28B* was also observed at the protein level, while *DPPA4* did not seem to change with passaging (Figure 6A-6C). In order to determine if continued passaging brought PSC-NPCs globally closer to 16-week-old fetal spinal cord-derived NPCs, we performed two Pearson correlations comparing the transcriptomes of the indicated samples. When using only those probe sets identified by the original *t*-test (Figure 3A; 2 769 less 46 probes that did not make it past the new filtering), we observed a small increase in similarity between PSC-NPCs and fetal-derived NPCs (Figure 6E). In addition, we observed that a small, but statistically significant, number of the original 2 723 probe sets were “corrected” upon extended passaging, including *LIN28A* and *LIN28B* (Figure 6F). Ultimately, when all probe sets were considered, we observed only a modest increase in global transcriptome similarity with extended passaging (Figure 6D), suggesting that simply culturing PSC-NPCs does not generate cells that are equivalent to their tissue-derived counterparts.

To functionally determine whether passaging PSC-NPCs promotes their developmental maturity, we compared the differentiation potential of PSC-NPCs before and after continued culturing. This analysis indicated that culturing PSC-NPCs for an additional month did increase their gliogenic capacity from < 1% to ~15% (Figure 6G), but not to levels typical of any tissue-derived NPCs we have tested to date (~50% - 80% gliogenic with NPCs derived from fetal tissue at 6.5-19 weeks of development (Figure 1A” and data not shown). These data further suggest that while continued culture can bring PSC-NPCs closer to their tissue-derived counterparts, this effect is small, and by itself is not sufficient to generate cells equivalent to tissue-derived NPCs. In addition, continued culture of PSC-NPCs in these conditions beyond 2 months (passage 4) led to the subsequent loss of NPC markers and differentiation capacity (data not shown). This suggests that, under these conditions, one cannot simply passage the cells indefinitely and expect continued developmental maturation *in vitro*.

To determine if PSC-NPCs would more closely resemble an NPC type isolated from younger fetal spinal

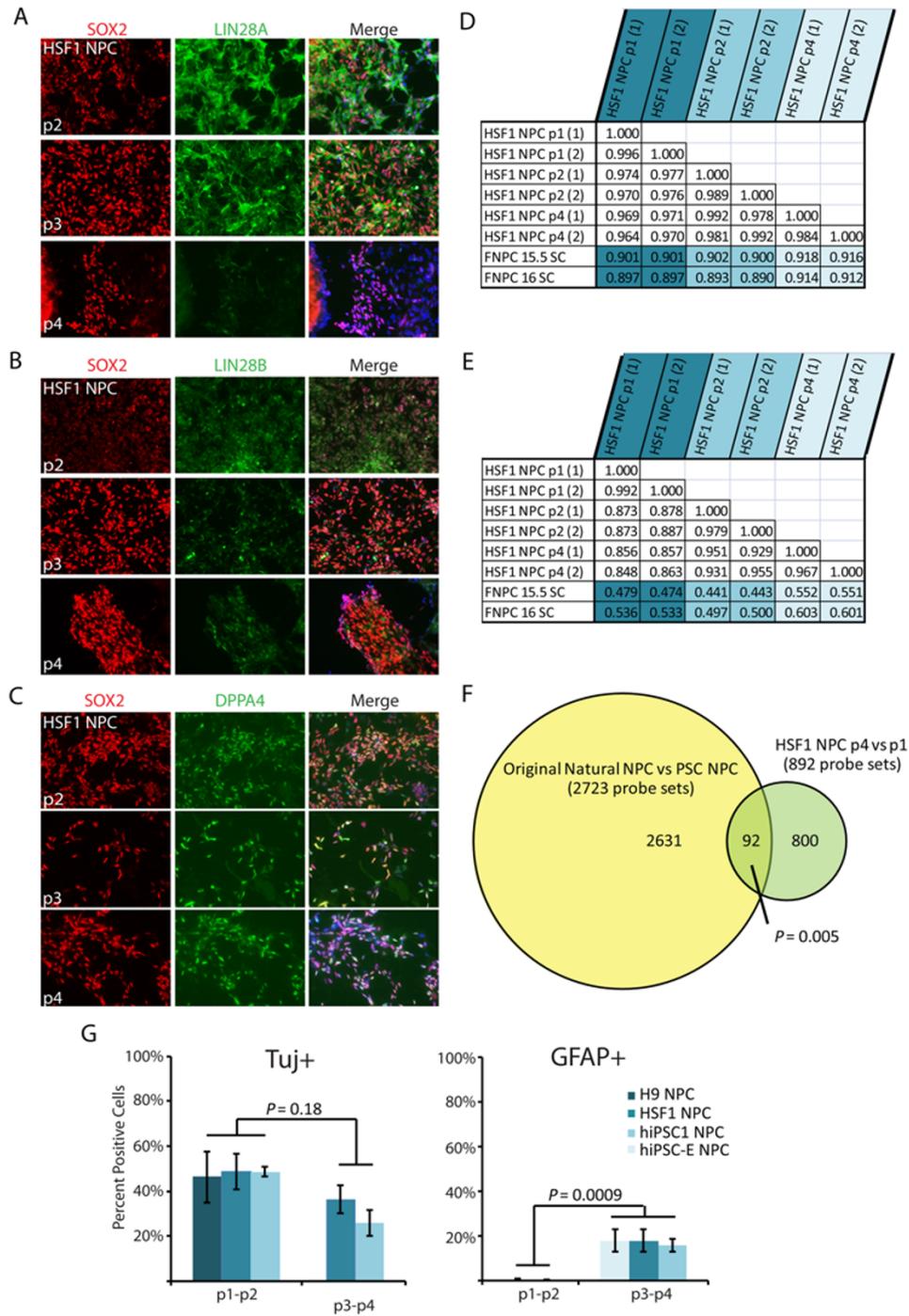


Figure 6 Continued passaging of PSC-NPCs reduces LIN28 expression and corrects a small portion of the gene expression discrepancies. **(A-C)** Immunofluorescence staining of HSF1-derived NPCs over four passages. Each passage represents ~5-7 days in culture. **(D)** Pearson correlation comparing global gene expression between HSF1 NPCs over several passages and NPCs derived from 16-week-old fetal spinal cord. **(E)** Pearson correlation including only those probe sets identified as different between PSC-NPC and Nat-NPCs (analysis from Figure 3A). **(F)** Venn diagram demonstrating the original differences identified in Figure 3A overlap significantly with gene expression differences between p1 and p4 PSC-NPCs (t -test $P < 0.01$; fold change ≥ 1.54). Direction of differential expression was taken into account. Statistical analysis performed by hypergeometric distribution. Note: later analyses were performed by normalizing and filtering only samples of the neural lineage. As a result the original 2 769 probe sets identified by analysis in Figure 3A were reduced to 2 723. **(G)** Percent of PSC-NPCs at the indicated passage undergoing neuronal (Tuj1) or glial (GFAP) differentiation following 3 weeks of differentiation.

cord, we profiled additional fetal samples from 6.5 to 8 weeks of development. Pearson correlation of the global transcriptome demonstrated a dramatic increase in similarity between PSC-derived NPCs and NPCs from young fetal spinal cord compared to 16-week-old spinal cord (Figure 7A). Ideally, a comparison between cells derived from fetal tissue earlier than 6 weeks would determine whether PSC derivatives accurately reflect their counterparts found during very early fetal development, but because of the lack of access to such tissue, this question is difficult to answer. However, data on gene expression across whole human embryos from 3-5 weeks of development recently became available [42].

We identified a list of 1 645 probe sets differentially expressed between 3-week (stage 9) and 5-week (stage 14) embryos. We overlaid these identified probe sets with the list of probe sets differentially expressed between PSC derivatives and their natural counterparts (Figure 3A; 105 is reduced to 36 due to the more restrictive chip used by Fang *et al.* See Materials and Methods for details) and found significant overlap. *LIN28A* was the most differentially expressed probe set between stage 9 and stage 14 embryos. Knowing that *LIN28A* was differentially expressed in both the NPC and hepatocyte lineages, but not the FBs of our data, we repeated the analysis and excluded FBs (Figure 7C). Amongst the 53 probe sets conserved in our data and the Fang *et al.* [42] data are *DPPA4* and *LIN28A* (note: *LIN28B* is not represented on the Fang *et al.* array). The expression pattern for the 46 unique genes represented by those 53 probe sets is represented by two heat maps, one including the samples from Fang *et al.* [42] and the other including our own samples (Figure 7D). The fact that a significant number of genes that are normally downregulated between 3-5 weeks of development appear to distinguish PSC derivatives from their tissue-derived counterparts further suggests that PSC derivatives might accurately recapitulate cells found prior to 6 weeks of development.

Discussion

Our data have revealed several important insights about differentiation from hPSCs. First, our data showed that hESCs and hiPSCs make specified derivatives that are nearly equivalent transcriptionally. This was surprising considering the vastly different circumstances by which hESCs and hiPSCs are derived, and in light of well-documented differences between them at the epigenetic and transcriptional levels in the undifferentiated state [2-6, 8, 9, 43, 44]. This similarity might be attributed to the fact that only high-passage hiPSCs (> 40) were used. However, even at high passage, a small number of

genes still appeared to distinguish the undifferentiated hiPSC and hESC lines used. The fact that these differences were largely undetectable in the differentiated state could suggest that the progeny of these cell types are more similar than their parental cells or that the genes differentially expressed between them are not expressed in the specified progeny.

Second, upon differentiation, we did not detect appreciable re-expression of the exogenous reprogramming factors. Because the expression of *OCT4* was undetectable in any PSC derivative, it is unlikely that the loci representing the retroviral reprogramming factors were re-activated upon differentiation or even continued culture. This finding could suggest that, at least in the contexts analyzed here, concerns over re-expression of oncogenic factors from hiPSC lines generated by viral integration could be mitigated by specification of hiPSCs to even a progenitor state, such as NPC, hepatoblast, or FB.

Third, both hESCs and hiPSCs made progeny that continued to express a group of genes known to play roles in very early embryonic development. While the progeny of ESCs have been proposed to represent embryonic cell types because of the primitive nature of the starting cell types [23, 25, 45], to our knowledge human pluripotent cell-derived progeny have not yet been placed into such a narrow developmental context. The most logical interpretation of our data is that the PSC derivatives generated here represent cells similar to those found earlier than 6 weeks of development.

It remains possible, however, that current protocols to make hPSC derivatives generate cells that do not completely represent cells found in tissue at any developmental stage. With limited access to tissues representing the earliest stages of development, this possibility will remain unexplored for now. It is also possible that new culture conditions could be defined that improve differentiation *in vitro* to better recapitulate that which occurs *in vivo*. In addition, it is possible that experimentally manipulating the expression of early embryonic genes described here could be used to accelerate development *in vitro*. Regardless of the fact that PSC derivatives produced in other labs also express some of the same early embryonic genes suggest that many protocols lead to a consistent result, so perhaps a shift in differentiation strategy or expectations is required.

Our data also suggest that simple approaches, such as continued culture, can further the development of PSC-derived cells, though this method also has its limitations as described above. Nevertheless, it is tempting to speculate that, upon specification, the differentiation process is more or less pre-determined by mechanisms that lock cells into a process that takes a specific amount of time,

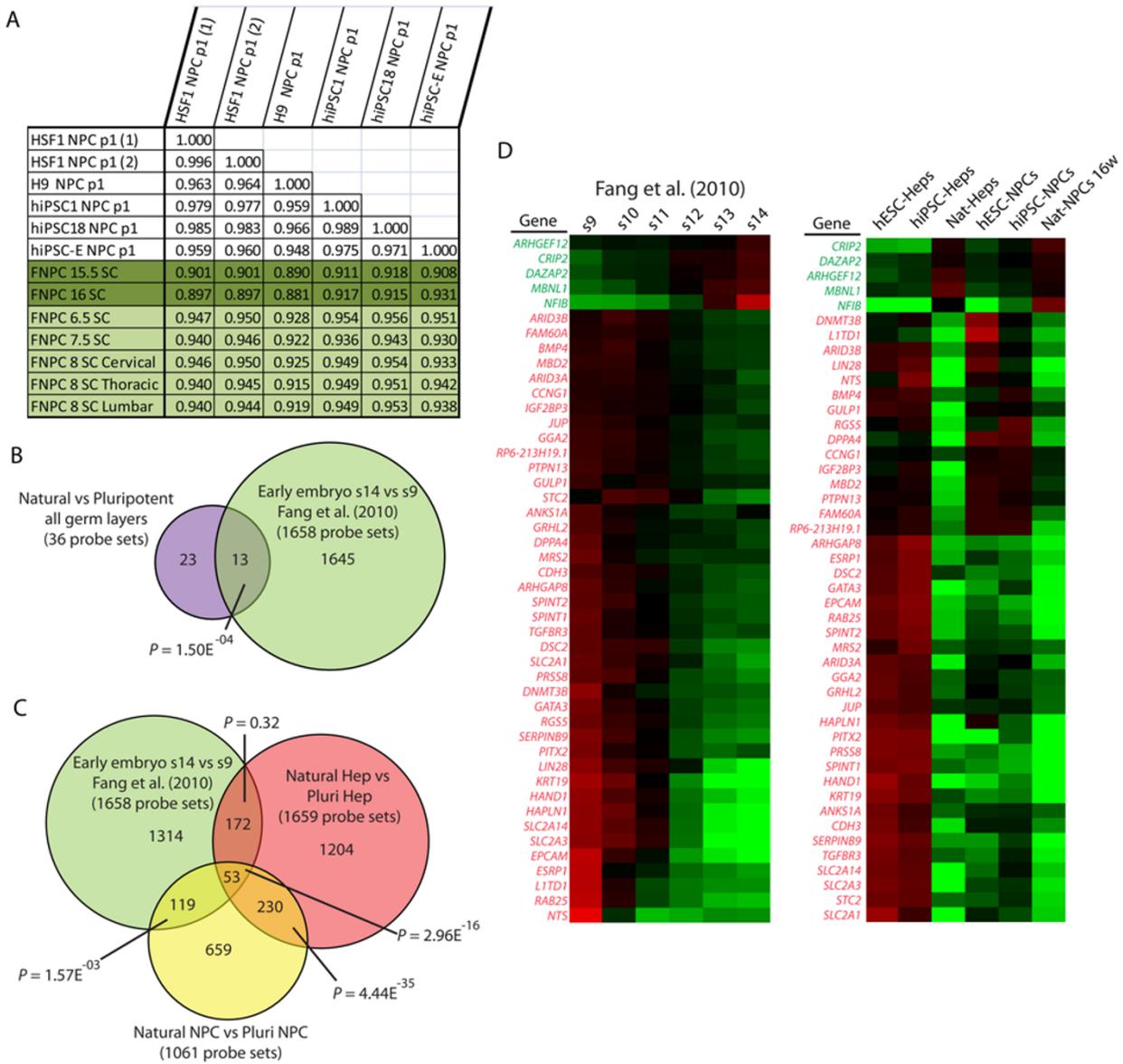


Figure 7 Evidence that PSC derivatives reflect cell types found prior to 6 weeks of development. **(A)** Pearson correlation comparing global gene expression between PSC-NPCs and fetal spinal cord NPCs. **(B)** Comparison of the original probe sets identified as different between PSC derivatives and their natural counterpart (Figure 3A) and those differentially expressed between stage 9 and stage 14 embryos (Fang *et al.* [42], *t*-test $P < 0.01$, fold change ≥ 1.54). **(C)** Venn diagram comparing the probe sets different between PSC derivatives and their tissue-derived counterparts for the Hep and NPC lineages and those differentially expressed between stage 9 and stage 14 embryos. **(D)** Heat maps generated for the 46 unique genes represented by the 53 probe sets shown in **C**. Left, samples include the six stages of embryonic development represented in Fang *et al.* [42]. Right, samples include PSC-derived NPCs and Heps and their respective natural counterparts. Green probe sets represent those genes upregulated over the course of development, while red probe sets are those downregulated over the course of development.

or number of cell divisions complete, regardless of the culture conditions employed. This idea has significant support from studies with *in vitro* murine development [45], where differentiation of mESCs under minimal

conditions allowed appropriate temporal and regional specification of neural tissue.

Recent work suggested that hPSC derivatives share hallmarks of gene expression with oncogenic cells [46].

Our data suggest that the presence of oncogenic hallmarks could be explained by persistent expression of genes in hPSC derivatives that are typical of the early embryo and known to be re-expressed in cancers. For example, while *LIN28* is not normally expressed in post-natal mammalian tissue, it was recently shown to be re-expressed in 15% of human cancers; and cancers with poor prognosis that are high in *LIN28* expression are low in *let-7* family expression [39]. This study also showed that overexpression of *LIN28* drove transformation of FBs, consistent with its proposed role in reprogramming [35, 38]. Therefore, if the derivatives of PSCs are to be used clinically, it could be important to take into account the residual expression of the early embryonic genes, particularly, *LIN28*. Finally, one of the great benefits of iPS technology is the ability to model human diseases *in vitro* using patient-derived cells. Our data would suggest that it could be difficult to model human diseases in this context, unless a phenotype manifests very early in development.

Materials and Methods

Human tissue sourcing

Primary human hepatocytes and FBs were acquired from Lonza (Switzerland). Fetal tissues were generated from discarded anonymized material obtained from elective terminations of first and second trimester pregnancies performed by the Family Planning Associates or the UCLA Medical Center. The fetal ages discussed throughout the paper represent the developmental age established 2 weeks after the first day of the last menstrual cycle. Tissues were harvested directly into sterile containers with PBS and transported on ice in PBS containing 5% FBS (Hyclone), 0.1% Ciprofloxacin HCl (10 µg/ml, Sigma), 1% amphotericin B (250 µg/ml, Invitrogen) and 1% penicillin-streptomycin (10 000 U/ml-10 000 µg/ml, Gibco) and processed the same day. Fetal spinal cord NPC lines were generated from 6-16-week-old spinal cords treated with collagenase and dispase for 1 h at 37 °C and plated on poly-ornithine/laminin-coated plates.

Cell culture

hESC and hiPSC were cultured as described previously [2, 47] in accordance with the UCLA ESCRO. Neural rosette derivation, NPC purification, and further differentiation to neurons and glia were performed as described [18]. Hepatocytes were differentiated using the four-stage protocol as described [20], with the exception of using FBS during the first 3 days of endoderm derivation. The AFP-GFP reporter [48] was transfected during differentiation using Lipofectamine 2000 (Invitrogen) and labeled cells were isolated using FACSA RIA (BD Biosciences). To generate FBs, EBs were cultured in adherent conditions using standard FB culturing media [47] and passaged until adopting typical FB morphology.

Immunostaining and western blot

Tissue acquired from embryonic spinal cord was fixed in 4% PFA for 1 followed by cryopreservation with sucrose and embed-

ding in O.C.T. Compound (Sakura). Fetal livers were freshly embedded in O.C.T. Compound. Both tissues were sectioned on a CM3050S cryostat (Leica) at 6-10 µm thickness. Liver sections were fixed in formalin for 10 min at room temperature prior to immunostaining. Immunostaining was performed as described [18, 50]. Coverslips were fixed in 4% PFA, blocked for 1 h in 10% serum + 0.1% Tween 20 (or 0.1% Triton-X-100), then incubated overnight at 4 °C with primary antibodies. Following primary antibody incubation, the coverslips/cryosections were incubated with Alexa Fluor secondary antibodies (Invitrogen) at room temperature for 1 h and mounted in Prolong Gold with DAPI (Invitrogen). All imaging was performed on Zeiss Axio Imager A1. Antibodies used include the following: mouse anti-DPPA4 (Abnova), goat anti-ALB (Bethyl Laboratories), rabbit anti-SOX2 (Biolegend), rabbit anti-LIN28A, rabbit anti-LIN28B, and rabbit anti-Nanog (Cell Signaling Technology), rabbit anti-PAX6, rabbit anti-Tuj1 (Covance), rabbit anti-GFAP (DAKO), anti-CD44 (Developmental Studies Hybridoma Bank), mouse anti-NESTIN (NeuroMics), goat anti-COL3A1, mouse anti-AFP, mouse anti-OCT3/4 (Santa Cruz Biotechnology Inc.), rabbit anti-SERPINA1 (Sigma Aldrich), and rabbit anti-Ki67 (Abcam). Western blot analysis was performed using standard procedures as described [50].

Assays for function

For periodic acid-schiff assay, cells were fixed with formalin-ethanol fixative and stained using standard protocol described in the manufacturer's manual (Sigma Aldrich). For albumin ELISA assay, media were collected from confluent wells and subjected to ELISA assay according to the manufacturer's protocol (Bethyl Laboratories). For Alizarin Red staining, natural and PSC-FBs were subjected to osteogenic induction and were stained with Alizarin Red as described [51]. Student's *t*-tests for the ELISA and NPC differentiation assays were performed in Excel.

Reporter assay

Cells transfected with the psiCHECK2- *let-7* 8× luciferase reporter (Addgene #20932) or psiCHECK2 control reporter (Promega) were lysed 72 h post-transfection and subjected to dual-glo luciferase assay as described in the manufacturer's protocol (Promega). The Renilla luciferase gene was driven by T7 promoter and contained eight *let-7* targeting sequences in the 3' UTR, and Firefly luciferase driven by a constitutive promoter as a transfection control. Luciferase assays were carried out in a GloMax 96 Microplate Luminometer (Promega).

Expression analysis

RNA isolation, reverse transcription, and real-time PCR were performed as described (Lowry *et al.* [47]). Microarray profiling was performed with Affymetrix Human HG-U133 2.0 Plus arrays as described [47, 52]. Data were normalized with Robust Multi-chip Algorithm in Genespring. Probe sets that were not expressed at a raw value of > 50 in at least 10% of samples were eliminated from further analysis. Note that outside data were normalized and analyzed separately, but with identical methods described above. Later, analysis on the NPC lineage was performed after separate normalization and filtering of only relevant samples. Outside data sets were collected from the Gene Expression Omnibus (NCBI) and includes: GSE19735 (Rafii), comparing hESC-derived endothelial cell (EC) to human umbilical vein ECs (HUVECs) and

human smooth muscle cells; GSE14897 (Duncan), comparing undifferentiated hESCs and hiPSCs to hepatocytes made from each; GSE20013 (Wilson), comparing ECs made from hESCs to ECs made from tissue (HUVEC); and GSE18887 (Fang), comparing entire human embryo samples from 3-5 weeks of development. All outside data were analyzed separately, but by employing the same methods. Fang *et al.* [42] utilized a U133 chip, which had fewer probe sets than our U133 2.0 plus chip. Only those probe sets that made it past the filtering on both chips were included in this analysis. Hierarchical Euclidian clustering was performed with complete linkage. Gene expression differences were judged to be significant if the *P*-value of the fold change was < 0.01 and at least 1.54-fold different between indicated samples. Gene expression differences were also judged by Benjamini-Hochberg correction for FDR at a *P*-value of < 0.05 and 1.54-fold change (Supplementary information, Figure S2). Further statistical analysis for hypergeometric distribution and three-way simulation was performed with R, package 2.9.2 as described [12]. Heat maps were generated by averaging the raw value of like samples and representing it as a ratio of the average of the raw value of all samples (Cluster 3.0, Java TreeView). Pearson correlations were generated in Excel.

Acknowledgments

We are grateful to Eric Wexler and Heather Martin for assistance in the acquisition of fetal tissues; to Margaret Baron for sharing the AFP-GFP reporter; to Soheila Azghadi, Hung Trinh, Otaren Aimuwu, Kimberly Loo and Adelaja Akinlolou for technical support; and to Mark Chin and Raj Sasidharan for assistance with bioinformatic analyses. We would like to acknowledge the core facilities at the UCLA including: FACS (Felicia Cordea and Jessica Scholes); the Clinical Microarray Core (Xinmin Li and Jian Zhou); and the hiPSC and hESC Cores. MP was supported by the Training Program in Genetic Mechanisms (GM007104-33). MP and DC were supported by Predoctoral Training Grants from CIRM at EEBSCRC (TG2-01169). This work was supported by a seed grant from CIRM (RS1-00259-1), a Basil O'Connor Starter Scholar Award (5-FY09-61), and the Fuller Foundation at EEB-SCRC (WEL). WEL holds the Maria Rowena Ross Chair in Cell Biology and Biochemistry.

References

- 1 Takahashi K, Tanabe K, Ohnuki M, *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; **131**:861-872.
- 2 Chin MH, Mason MJ, Xie W, *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 2009; **5**:111-123.
- 3 Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS ONE* 2010; **5**:e8975.
- 4 Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR. Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS ONE* 2009; **4**:e7076.
- 5 Bock C, Kiskinis E, Verstappen G, *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 2011; **144**:439-452.
- 6 Lister R, Pelizzola M, Kida YS, *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 2011; **471**:68-73.
- 7 Narsinh KH, Sun N, Sanchez-Freire V, *et al.* Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. *J Clin Invest* 2011; **121**:1217-1221.
- 8 Polo JM, Liu S, Figueroa ME, *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol* 2010; **28**:848-855.
- 9 Kim K, Doi A, Wen B, *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* 2010; **467**:285-290.
- 10 Ohi Y, Qin H, Hong C, *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat Cell Biol* 2011; **13**:541-549.
- 11 Bar-Nur O, Russ HA, Efrat S, Benvenisty N. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet Beta cells. *Cell Stem Cell* 2011; **9**:17-23.
- 12 Chin MH, Pellegrini M, Plath K, Lowry WE. Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. *Cell Stem Cell* 2010; **7**:263-269.
- 13 Osafune K, Caron L, Borowiak M, *et al.* Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat Biotechnol* 2008; **26**:313-315.
- 14 Feng Q, Lu SJ, Klimanskaya I, *et al.* Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells* 2010; **28**:704-712.
- 15 Boulting GL, Kiskinis E, Croft GF, *et al.* A functionally characterized test set of human induced pluripotent stem cells. *Nat Biotechnol* 2011; **29**:279-286.
- 16 Si-Tayeb K, Noto FK, Nagaoka M, *et al.* Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* 2010; **51**:297-305.
- 17 Pfannkuche K, Liang H, Hannes T, *et al.* Cardiac myocytes derived from murine reprogrammed fibroblasts: intact hormonal regulation, cardiac ion channel expression and development of contractility. *Cell Physiol Biochem* 2009; **24**:73-86.
- 18 Karumbayaram S, Novitch BG, Patterson M, *et al.* Directed differentiation of human-induced pluripotent stem cells generates active motor neurons. *Stem Cells* 2009; **27**:806-811.
- 19 Zhang J, Wilson GF, Soerens AG, *et al.* Functional cardiomyocytes derived from human induced pluripotent stem cells. *Circ Res* 2009; **104**:e30-41.
- 20 Song Z, Cai J, Liu Y, *et al.* Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells. *Cell Res* 2009; **19**:1233-1242.
- 21 Choi KD, Yu J, Smuga-Otto K, *et al.* Hematopoietic and endothelial differentiation of human induced pluripotent stem cells. *Stem Cells* 2009; **27**:559-567.
- 22 Kennedy M, D'Souza SL, Lynch-Kattman M, Schwantz S, Keller G. Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood* 2007; **109**:2679-2687.
- 23 Keller G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* 2005; **19**:1129-1155.
- 24 Pankratz MT, Li XJ, Lavaute TM, Lyons EA, Chen X, Zhang

- SC. Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. *Stem Cells* 2007; **25**:1511-1520.
- 25 Elkabetz Y, Panagiotakos G, Al Shamy G, Socci ND, Tabar V, Studer L. Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev* 2008; **22**:152-165.
- 26 Deneen B, Ho R, Lukaszewicz A, Hochstim CJ, Gronostajski RM, Anderson DJ. The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron* 2006; **52**:953-968.
- 27 Viswanathan SR, Daley GQ. Lin28: A microRNA regulator with a macro role. *Cell* 2010; **140**:445-449.
- 28 Madan B, Madan V, Weber O, *et al.* The pluripotency-associated gene *Dppa4* is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis. *Mol Cell Biol* 2009; **29**:3186-3203.
- 29 Moss EG, Tang L. Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev Biol* 2003; **258**:432-442.
- 30 Zhu H, Shah S, Shyh-Chang N, *et al.* Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat Genet* 2010; **42**:626-630.
- 31 West JA, Viswanathan SR, Yabuuchi A, *et al.* A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* 2009; **460**:909-913.
- 32 Merrill BJ, Pasolli HA, Polak L, *et al.* Tcf3: a transcriptional regulator of axis induction in the early embryo. *Development* 2004; **131**:263-274.
- 33 Pereira L, Yi F, Merrill BJ. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol* 2006; **26**:7479-7491.
- 34 Ambros V, Horvitz HR. Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* 1984; **226**:409-416.
- 35 Yu J, Vodyanik MA, Smuga-Otto K, *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; **318**:1917-1920.
- 36 Lehrbach NJ, Armisen J, Lightfoot HL, *et al.* LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nat Struct Mol Biol* 2009; **16**:1016-1020.
- 37 Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* 2009; **16**:1021-1025.
- 38 Heo I, Joo C, Kim YK, *et al.* TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* 2009; **138**:696-708.
- 39 Viswanathan SR, Powers JT, Einhorn W, *et al.* Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet* 2009; **41**:843-848.
- 40 Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* 2008; **32**:276-284.
- 41 Iwasaki S, Kawamata T, Tomari Y. *Drosophila* argonaute1 and argonaute2 employ distinct mechanisms for translational repression. *Mol Cell* 2009; **34**:58-67.
- 42 Fang H, Yang Y, Li C, *et al.* Transcriptome analysis of early organogenesis in human embryos. *Dev Cell* 2010; **19**:174-184.
- 43 Doi A, Park IH, Wen B, *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 2009; **41**:1350-1353.
- 44 Hawkins RD, Hon GC, Lee LK, *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 2010; **6**:479-491.
- 45 Gaspard N, Bouchet T, Hourez R, *et al.* An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* 2008; **455**:351-357.
- 46 Ghosh Z, Huang M, Hu S, Wilson KD, Dey D, Wu JC. Dissecting the oncogenic potential of human embryonic and induced pluripotent stem cell derivatives. *Cancer Res* 2011; **71**:5030-5039.
- 47 Lowry WE, Richter L, Yachechko R, *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci USA* 2008; **105**:2883-2888.
- 48 Kwon GS, Fraser ST, Eakin GS, *et al.* Tg(Afp-GFP) expression marks primitive and definitive endoderm lineages during mouse development. *Dev Dyn* 2006; **235**:2549-2558.
- 49 Lee R, Chang SY, Trinh H, *et al.* Genetic studies on the functional relevance of the protein prenyltransferases in skin keratinocytes. *Hum Mol Genet* 2010; **19**:1603-1617.
- 50 Lowry WE, Blanpain C, Nowak JA, Guasch G, Lewis L, Fuchs E. Defining the impact of beta-catenin/Tcf transactivation on epithelial stem cells. *Genes Dev* 2005; **19**:1596-1611.
- 51 Tang Z, Sahu SN, Khadeer MA, Bai G, Franklin RB, Gupta A. Overexpression of the ZIP1 zinc transporter induces an osteogenic phenotype in mesenchymal stem cells. *Bone* 2006; **38**:181-198.
- 52 Tchieu J, Kuoy E, Chin MH, *et al.* Female human iPS cells retain an inactive X-chromosome. *Cell Stem Cell* 2010; **7**:329-342.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)