

Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming

Xiaodong Liu^{1-3,11}, Christian M Nefzger^{1-3,11}, Fernando J Rossello¹⁻³ , Joseph Chen¹⁻³, Anja S Knaupp¹⁻³, Jaber Firas¹⁻³, Ethan Ford^{4,5}, Jahnvi Pflueger^{4,5}, Jacob M Paynter¹⁻³, Hun S Chy^{3,6}, Carmel M O'Brien^{3,6}, Cheng Huang⁷, Ketan Mishra¹⁻³, Margeaux Hodgson-Garms¹⁻³, Natasha Jansz^{8,9}, Sarah M Williams^{1-3,10}, Marnie E Blewitt^{8,9}, Susan K Nilsson^{3,6}, Ralf B Schittenhelm⁷, Andrew L Laslett^{3,6}, Ryan Lister^{4,5} & Jose M Polo¹⁻³ 

Recent reports on the characteristics of naive human pluripotent stem cells (hPSCs) obtained using independent methods differ. Naive hPSCs have been mainly derived by conversion from primed hPSCs or by direct derivation from human embryos rather than by somatic cell reprogramming. To provide an unbiased molecular and functional reference, we derived genetically matched naive hPSCs by direct reprogramming of fibroblasts and by primed-to-naive conversion using different naive conditions (NHSM, RSeT, 5iLAF and t2iLGöY). Our results show that hPSCs obtained in these different conditions display a spectrum of naive characteristics. Furthermore, our characterization identifies *KLF4* as sufficient for conversion of primed hPSCs into naive t2iLGöY hPSCs, underscoring the role that reprogramming factors can play for the derivation of bona fide naive hPSCs.

Human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) are typically maintained in a 'primed pluripotent state' similar to the state in which mouse epiblast stem cells (EpiSCs) are maintained¹. Naive pluripotency represents an earlier stage of development that differs from the primed state in signaling state, metabolic requirement, colony morphology, single-cell cloning efficiency, transcriptional and epigenetic profiles, and chimera formation. Multiple protocols have been published to derive and maintain naive hPSCs²⁻⁷ that exhibit some characteristics of human inner cell mass (ICM) cells, but different characteristics have often been reported in each study. This has complicated the identification of a consensus standard for characteristics possessed by naive hPSCs^{4,5,7-10}. Furthermore, it has been reported that the naive state (characterized by two

active X chromosomes in female lines) can reset common abnormalities of X chromosome inactivation found in female primed hPSCs^{11,12} enabling proper *de novo* X inactivation upon differentiation¹¹. Importantly, these different characteristics could affect downstream applications in basic research and clinical translation (reviewed in ref. 13). Published protocols have primarily focused on conversion from primed cells²⁻⁶ or direct derivation from human embryos^{3,5-7,9}. Here, using five distinct methodologies and genetically matched cells, we investigated whether fibroblasts could be directly reprogrammed into naive-like hiPSCs and compared these cell lines molecularly and functionally.

RESULTS

Reprogramming of human fibroblasts in naive culture conditions

We first converted primed hiPSCs (derived from two donors, 32F and 55F) in NHSM, 5iLAF^{3,5}, t2iLGöY⁷ and RSeT conditions (**Supplementary Fig. 1a** and **Supplementary Table 1**) (RSeT is a commercial medium based on NHSM). As previously reported, we could derive naive-like cells by exposing primed cells to NHSM, RSeT or 5iLAF conditions (termed cNHSM, cRSeT and c5iLAF); but the t2iLGöY condition did not produce naive hiPSCs since *NANOG* and *KLF2* expression are required for conversion⁴. We observed round, dome-shaped colonies in RSeT and 5iLAF conditions, in contrast to flat colonies in the primed condition and colonies with an intermediate phenotype in the NHSM condition (**Supplementary Fig. 1b**). Immunostaining showed that all cell lines were positive for pluripotency markers TRA-1-60, *NANOG* and *OCT4*, consistent with previous reports^{3,5} (**Supplementary Fig. 1c**). As reported for mouse naive

¹Department of Anatomy and Developmental Biology, Monash University, Wellington Road, Clayton, Victoria, Australia. ²Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Wellington Road, Clayton, Victoria, Australia. ³Australian Regenerative Medicine Institute, Monash University, Wellington Road, Clayton, Victoria, Australia. ⁴ARC Center of Excellence in Plant Energy Biology, The University of Western Australia, Perth, Western Australia, Australia. ⁵The Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia. ⁶Manufacturing, CSIRO, Clayton, Victoria, Australia. ⁷Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia. ⁸The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia. ⁹University of Melbourne, Melbourne, Victoria, Australia. ¹⁰Monash Bioinformatics Platform, Monash University, Wellington Road, Clayton, Victoria, Australia. ¹¹These authors contributed equally to this work. Correspondence should be addressed to J.M.P. (jose.polo@monash.edu).

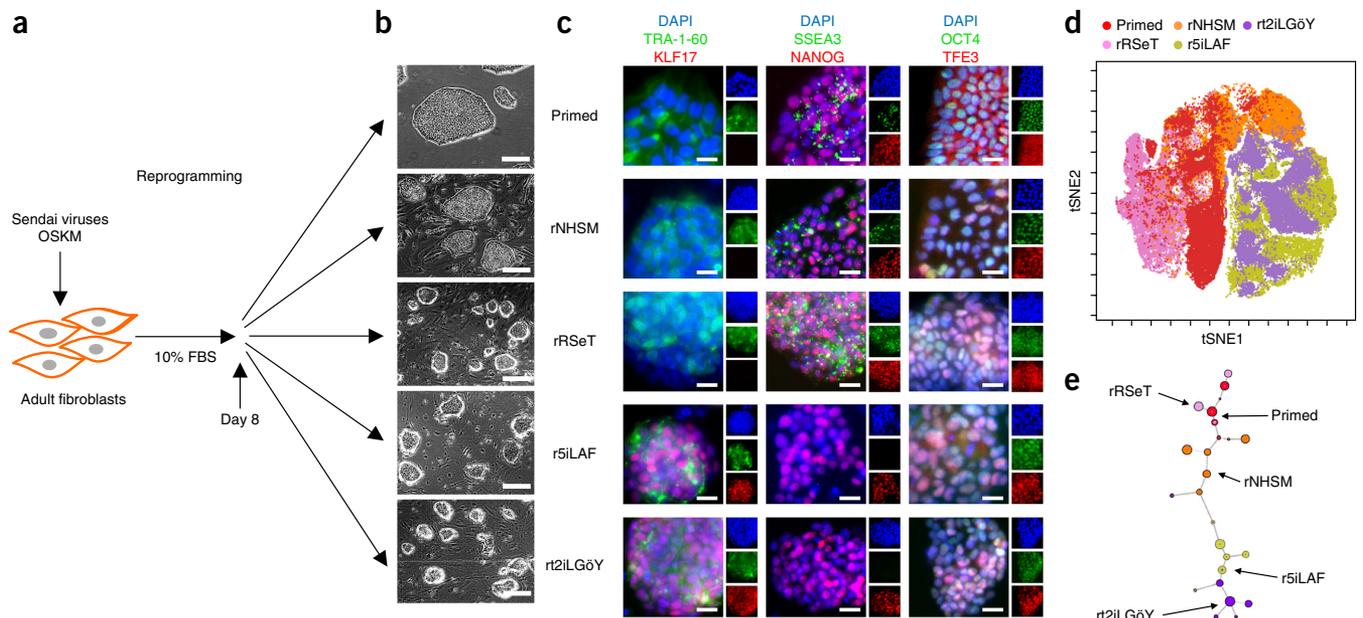


Figure 1 | Direct reprogramming of human fibroblasts into naive hiPSCs under different conditions. **(a)** Schematic overview for direct generation of different types of naive hiPSCs via somatic cell reprogramming using the Sendai viral system and indicated culture conditions. **(b)** Resulting colony morphologies (scale bars, 250 μ m). **(c)** Immunofluorescence staining for pluripotency markers TRA-1-60, SSEA3, NANOG and OCT4 and naive-associated markers KLF17 and TFE3 (scale bars, 25 μ m). **(d,e)** Multidimensional representation of flow cytometry data for primed and different types of naive hiPSCs by **(d)** viSNE map and **(e)** SPADE tree.

ESCs¹⁴, cNHSM, cRSeT and c5iLAF cells (but not primed lines) exhibited nuclear enrichment for TFE3 (**Supplementary Fig. 1c**). Furthermore, only c5iLAF cells expressed KLF17, a naive-associated marker and lacked expression of SSEA3 (**Supplementary Fig. 1c**). Conversely, primed, cNHSM and cRSeT cells were positive for SSEA3 and negative for KLF17.

Next, we attempted to reprogram human fibroblasts with Sendai viruses expressing *OCT4*, *SOX2*, *KLF4* and *c-MYC* (*OSKM*) under primed, NHSM, RSeT, 5iLAF and t2iLGöY conditions (**Fig. 1a** and **Supplementary Table 1**). We generated naive rNHSM, rRSeT and r5iLAF hiPSCs; and although it has not been previously reported, we also obtained naive hiPSCs in the t2iLGöY condition. These cells exhibited similar colony morphologies (**Fig. 1b**) and marker expression (**Fig. 1c**) to the corresponding converted naive cells (**Supplementary Fig. 1b,c**). Additionally, naive rt2iLGöY cells, like c5iLAF and r5iLAF cells, were positive for KLF17 and negative for SSEA3 (**Fig. 1c** and **Supplementary Fig. 1c**).

We further investigated a panel of cell-surface markers for primed and/or human naive cells^{9,15–17} and employed spanning-tree progression analysis of density-normalized events (SPADE) and viSNE^{18–20} to allow multidimensional separation and visualization of cell populations. Naive rRSeT cells and primed cells clustered closely in SPADE trees and viSNE maps, as they were positive for CD24, SSEA3, SSEA4, F11R, NLGN4X and TRA-1-60 (**Fig. 1d,e** and **Supplementary Fig. 2a**). In contrast, r5iLAF and rt2iLGöY cells clustered separately, and they were negative for CD24, SSEA3, SSEA4 and NLGN4X but TRA-1-60 low and F11R high (**Supplementary Fig. 2a,b**). Interestingly, naive rNHSM cells exhibited an intermediate phenotype (positive for CD24, SSEA4, F11R and TRA-1-60) while displaying low expression of SSEA3 and NLGN4X (**Supplementary Fig. 2a**). In summary, surface marker

profiling confirmed previous reports using CD24, SSEA3, SSEA4 and NLGN4X^{9,15–17} and revealed that high levels of F11R expression are characteristic of naive r5iLAF and rt2iLGöY cells, which distinguishes them from naive rNHSM, rRSeT and primed cells.

Naive hiPSCs obtained in different conditions represent a spectrum of naive characteristics whether derived by conversion or reprogramming

We performed whole-transcriptome profiling on our cohort of genetically matched naive and primed hiPSCs (from at least two different donors) derived via media conversion and reprogramming at early and late passages. Expression profiles showed that naive c5iLAF, r5iLAF and rt2iLGöY hiPSCs upregulated naive-associated genes and repressed primed-related genes (**Fig. 2a**). Upregulation of naive pluripotency-related genes in NHSM and RSeT naive hiPSCs was moderate (**Fig. 2a**). Hierarchical clustering (**Fig. 2b**) and multidimensional scaling (MDS) (**Supplementary Fig. 2c**) showed that naive c5iLAF, r5iLAF and rt2iLGöY hiPSCs clustered together, separately from primed cells. Interestingly, naive NHSM hiPSCs adopted an intermediate expression profile. Although hundreds of genes were differentially expressed between primed and naive cRSeT or rRSeT cells, they clustered closely to primed cells. Importantly, naive hiPSCs obtained by primed-to-naive conversion or reprogramming clustered together within each condition (**Fig. 2b**). We then compared these transcriptomes with previous studies^{3–5} and with a human ICM data set²¹. MDS revealed that our naive c5iLAF, r5iLAF and rt2iLGöY hiPSCs are transcriptionally analogous to reported naive hESCs cultured in 5iL/A or 6iL/A and t2iL+Gö conditions^{4,5} (**Fig. 2c**). Moreover, these populations clustered closest to cells from the human ICM²¹, which indicates, from a transcriptional viewpoint, that these naive

cells are more similar to the *in vivo* ground state counterpart. In addition, we observed that naive r5iLAF and rt2iLGöY hiPSCs exhibited a slower growth rate than that of primed cells (Supplementary Fig. 2d and Supplementary Table 2). In contrast, naive cRSeT and rRSeT cells, while distinct from primed cells, clustered apart from cells of the human ICM. Furthermore, our naive cNHSM and rNHSM cells clustered together with the originally derived NHSM hiPSCs and showed an intermediate expression pattern between primed and naive 5iLAF and t2iLGöY cells. Remarkably, the RNA-seq-based clustering is consistent with the marker-based SPADE and viSNE analyses, and this suggests the usefulness of the panel of cell-surface markers we identified to distinguish naive hiPSCs from primed cells.

Consistent with reports using converted naive hESCs^{4,9,10}, Methyl-Seq target enrichment bisulfite sequencing indicated a global DNA hypomethylation pattern in naive r5iLAF and rt2iLGöY hiPSCs (Fig. 2d), a phenomenon characteristic of human early blastocysts²². In contrast, naive rNHSM and rRSeT hiPSCs exhibited global DNA methylation levels similar to those of primed hiPSCs (Fig. 2d). Moreover, differentially methylated regions (DMRs) within imprinted loci, reported to be lost in 5iLAF and t2iL+Gö naive hESCs converted from primed cells^{9,10,23}, also showed loss of methylation in the naive r5iLAF and rt2iLGöY hiPSCs obtained by direct reprogramming, whereas the methylation of these regions was maintained in primed, naive rNHSM and rRSeT hiPSCs (Fig. 2e). Repriming of naive rt2iLGöY hiPSCs obtained by direct reprogramming reversed global DNA hypomethylation; however, the loss of primary imprinted DMRs was not re-established (Fig. 2d,e and Supplementary Fig. 2e), as was similarly observed for naive 5iLAF hESCs^{9,10}.

Genomic stability and differentiation potential of naive cells

The reported chromosomal instability in naive PSCs^{5,9} is a concern to the field in general. We analyzed our cells for up to 20 passages and confirmed karyotypic instability in naive c5iLAF and r5iLAF hiPSCs (Fig. 3a). However, the majority of other cell lines analyzed, both primed and naive (cNHSM, rNHSM, cRSeT, rRSeT and rt2iLGöY), exhibited normal karyotypes or a low degree of instability. We detected abnormalities in one of our rt2iLGöY cell lines (55F), but karyotypic analysis of the original fibroblasts revealed a pre-existing abnormality (Fig. 3a). Overall, in our hands, naive NHSM, RSeT and t2iLGöY hiPSCs can be maintained for 20 passages without gaining substantial chromosomal abnormalities, unlike naive 5iLAF hiPSCs.

We performed teratoma assays to evaluate the differentiation of naive hiPSCs. Naive r5iLAF and rt2iLGöY hiPSCs did not form teratomas (0/2 and 0/4, respectively) in comparison to primed (2/2), naive rNHSM (2/2) and rRSeT (2/2) hiPSCs (Fig. 3b). We do not exclude that naive r5iLAF and rt2iLGöY cells might produce teratomas at low frequency when more experiments are performed, as has been previously demonstrated^{4,5}. Nevertheless, our results indicate that naive r5iLAF and rt2iLGöY cells are less proficient at forming teratomas compared to the other conditions.

We also assessed the potential of cell lines to form embryoid bodies (EBs) and quantified trilineage potential using ScoreCard assays^{24,25}. All cell lines formed EBs (Fig. 3c), but only primed,

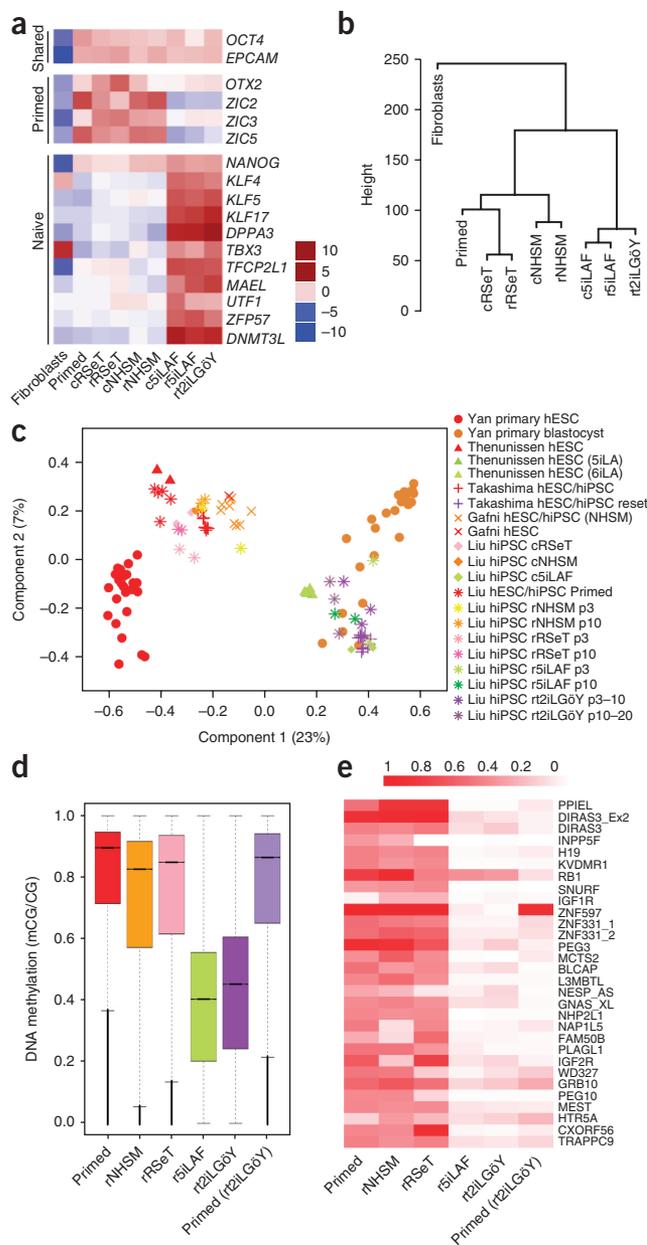


Figure 2 | Transcriptional and epigenetic characterization of naive hiPSCs. **(a)** Heatmap depicting expression levels of pan-pluripotent, primed and naive pluripotency-associated marker genes in media-converted and reprogrammed cell types. **(b)** Unsupervised hierarchical clustering for media-converted and reprogrammed cell types. Sample distances between samples was calculated using Euclidean distance and clustering performed using complete linkage method. Height is the distance metric between clusters calculated with the complete linkage method. **(c)** Multidimensional scaling analysis integrating primed and naive hiPSCs derived in this study with previously published data sets^{3-5,21}. **(a-c)** Two donors (32F and 55F) were used to generate all the hiPSCs except for cNHSM (one donor, 32F). Each dot represents a different cell culture sample per indicated condition. **(d)** Box plot of global DNA methylation status of reprogrammed primed and naive hiPSCs. The box plot shows 25/75th percentile, the line is the median value of all analyzed regions in one replicate, and the box plot whiskers extend to the most extreme point, 1.5 times the interquartile range from the box. **(e)** Heatmap of their DNA methylation status at primary imprinted loci. (For methylation studies, $n = 1$; donor 32F was used).

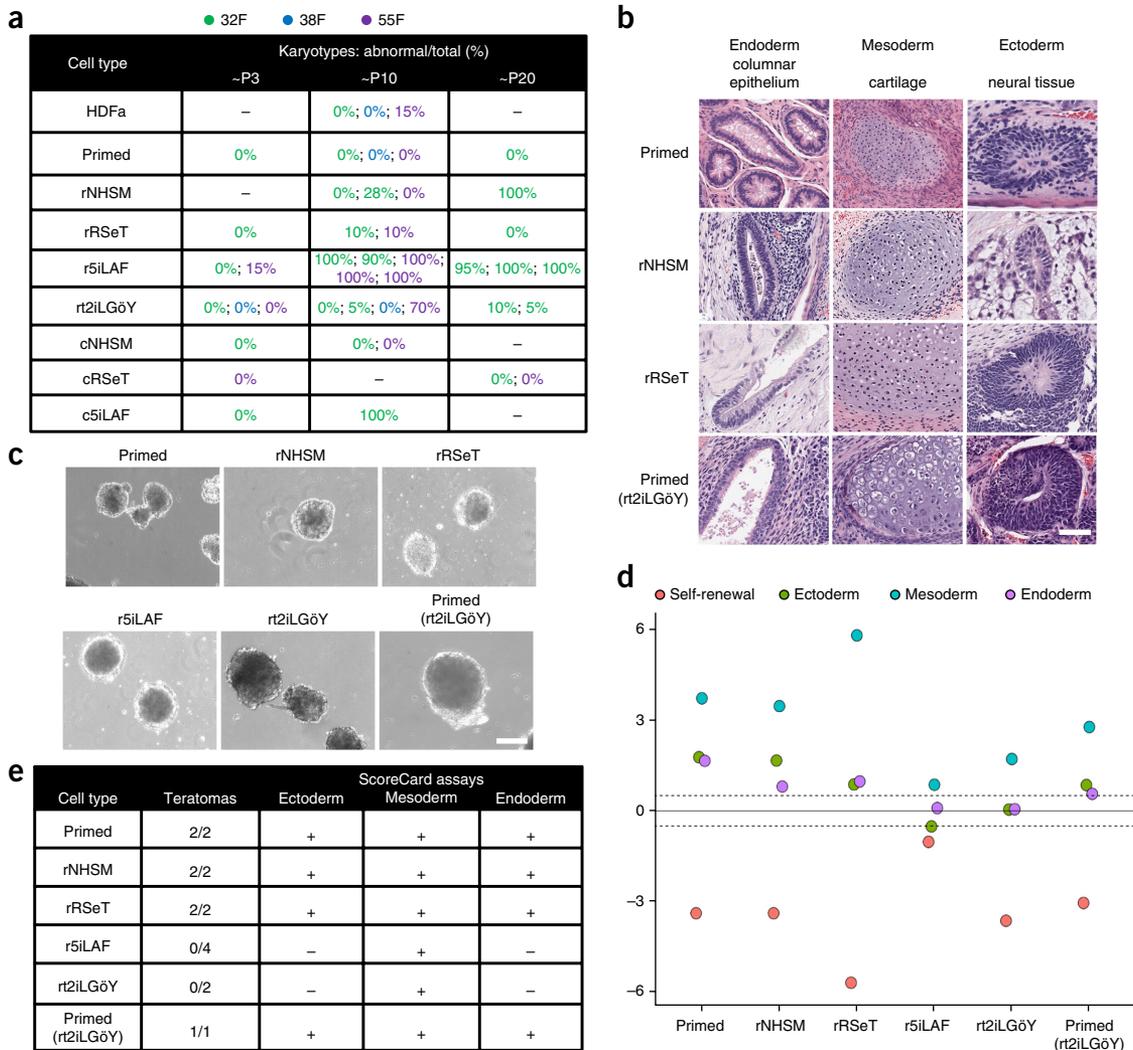


Figure 3 | Functional characterization of naive hiPSCs. **(a)** Table with karyotypes of cell lines derived in this study from three donors (32F in green, 38F in blue and 55F in purple) at different passage numbers, as indicated. **(b)** Teratomas generated after injection of primed, rNHSM, rRSeT and primed (rt2iLGöY) hiPSCs into the testis of NOD-SCID mice. Representative images for each germ layer are shown (scale bar, 200 μ m). **(c)** Bright-field images for embryoid bodies (scale bar, 100 μ m). **(d)** Dot plot of scores generated by the ScoreCard algorithm (32F background $n = 1$). **(e)** Overview table of results for Teratoma (ratios indicate the number of teratomas formation over the number of animals injected) and ScoreCard assays ('+' or '–' indicate germ layer contribution or lack thereof).

naive rNHSM and rRSeT hiPSCs showed trilineage differentiation (Fig. 3d,e). In contrast, naive r5iLAF and rt2iLGöY hiPSCs formed EBs that primarily expressed markers of differentiation into the mesoderm germ layer (Fig. 3d). We hypothesized that trilineage differentiation potential could be restored after pluripotency priming (Supplementary Fig. 2e). Our results showed that upon priming of naive rt2iLGöY hiPSCs, the cells could generate both teratomas and EBs with trilineage potential (Fig. 3b–e). In summary, naive rNHSM and rRSeT hiPSCs showed trilineage differentiation potential, whereas naive rt2iLGöY hiPSCs required pluripotency priming to give rise to all three germ layers, which possibly reflects a developmental requirement.

We further validated that naive hiPSCs generated by integration-free mRNA (unmodified non-self-replicating RNA molecules) in rt2iLGöY conditions appear to be highly similar, both morphologically and molecularly, to cells derived using the Sendai system (Supplementary Fig. 3).

Primed hPSCs conversion to naive hPSCs by *OSKM*

In previous studies, primed cells were converted to the t2iL+Gö naive state by the overexpression of *NANOG* and *KLF2* (ref. 4). We tested whether transient *OSKM* overexpression could also be employed to convert established human primed hPSC lines to the naive t2iLGöY state (Fig. 4a). Round dome-shaped colonies were observed after a few passages in wells transduced with *OSKM* (termed ct2iLGöY *OSKM*) (Fig. 4a). Naive cells regained *KLF17* and lost *SSEA3* immunoreactivity and exhibited enhanced nuclear localization of *TFE3* (Fig. 4b). Transcriptionally they clustered with c5iLAF, r5iLAF and rt2iLGöY hiPSCs and cells of the human ICM (Fig. 4c). Importantly, these cells were karyotypically stable after 20 passages (Fig. 4d).

Primed hPSCs conversion into naive hPSCs by *KLF4* alone

Given that naive rt2iLGöY hiPSCs are karyotypically stable and exhibit a transcriptome similar to that of human ICM cells, we

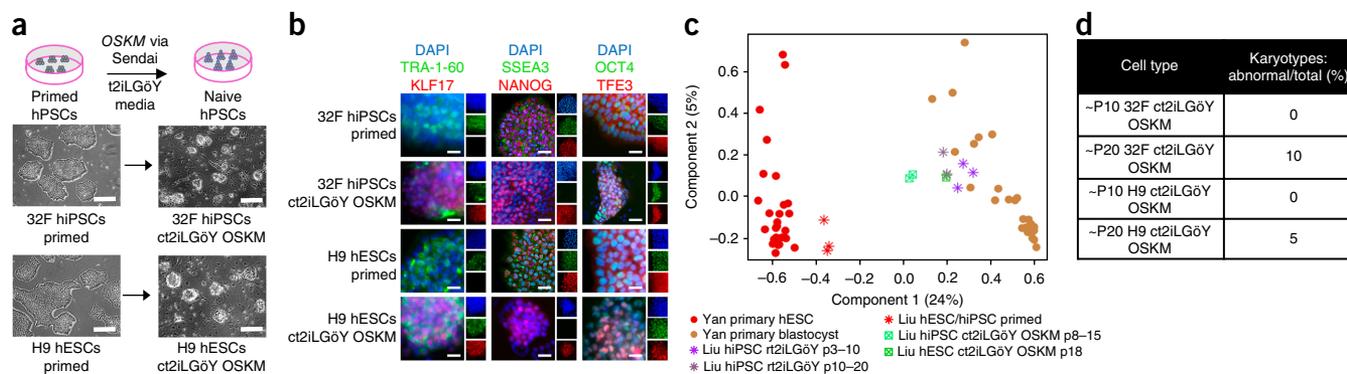


Figure 4 | Conversion of primed hPSCs to naive hPSCs by *OSKM*. (a) Schematic representation of experimental design for primed-to-naive conversion using *OSKM* via Sendai transduction in t2iLGöY condition (top panel) and bright-field images of cultures before and after transduction. Scale bar, 250 μ m (bottom panels). (b) Immunofluorescence labeling of established primed hiPSCs, hESCs, naive ct2iLGöY OSKM hiPSCs and ct2iLGöY OSKM hESCs. Scale bar, 25 μ m. (c) MDS plot for naive ct2iLGöY OSKM hiPSCs/hESCs ($n = 3$ cell donors: H9 ESC, 32F iPSC and 38F iPSC) as well as primed ($n = 3$ cell donors: H9 ESC, 32F iPSC and 55F iPSC) and naive rt2iLGöY hiPSCs ($n = 2$ donors: 32F and 55F) at the indicated passages integrated with a previously published human ICM data set²¹. Each dot represents a different cell culture sample per indicated condition. (d) Table with karyotypes of naive ct2iLGöY OSKM (two donors: H9 ESC and 32F iPSC at indicated passages).

investigated them further. We examined how the proteome of rt2iLGöY hiPSCs differs from that of primed cells (Fig. 5a and Supplementary Table 3) and found several enzymes indicative of high oxidative phosphorylation metabolism (citrate synthase (CS), isocitrate dehydrogenase 3 alpha (IDH3A), nicotinamide nucleotide transferase (NNT)) upregulated. Moreover, elevated levels of cell–cell adhesion proteins such as F11R, AHNK nucleoprotein (AHANK) and plectin (PLEC) were detected in naive rt2iLGöY hiPSCs, and these elevated levels were potentially associated with the dome-shaped colonies observed. The most significant protein groups expressed at lower levels in naive rt2iLGöY hiPSCs, compared with primed cells, were associated through gene ontology analysis with mitotic nuclear division. We also found a reduced level of DNMT3B but an elevated level of DNMT3L in naive rt2iLGöY hiPSCs (Fig. 5a), which is similar to the pattern found in naive 5iLAF hESCs⁹. Despite the fact that Takashima and colleagues demonstrated that conversion of primed hPSCs in the t2iL+Gö condition required the temporary expression of *NANOG* and *KLF2*⁴, we did not find *KLF2* to be expressed at the protein or RNA level in naive rt2iLGöY cells. A similar finding has been reported for cells of the human ICM^{21,26}. Intriguingly, we observed a high level of expression of endogenous *KLF4* at both the RNA and protein levels (Fig. 5a), which indicated that *KLF4* may play an important role in regulating human naive pluripotency. Additionally, we observed the highest levels of *KLF4* expression in cell types identified to cluster closely to human ICM (r5iLAF and rt2iLGöY). Conversely, *KLF4* expression levels were intermediate for rNHSM and rRSeT hiPSCs and lowest for primed hiPSCs (Fig. 5b and Supplementary Table 4). On this note, *Klf4* has been demonstrated to be crucial for the establishment of naive pluripotency in the mouse system^{27,28}.

We tested whether primed hESCs and hiPSCs could be reprogrammed to the naive state using just *KLF4* in the t2iLGöY condition (Fig. 5c). We observed round dome-shaped colonies in wells transduced with *KLF4* Sendai viruses (termed ct2iLGöY K) (Fig. 5c), and immunostaining demonstrated that ct2iLGöY K cells had gained *KLF17* and lost *SSEA3* immunoreactivity and exhibited enhanced nuclear localization of *TFE3*, which is typical of naive phenotypes (Supplementary Fig. 4a). ct2iLGöY K cells were karyotypically

stable for at least 20 passages (Supplementary Fig. 4b), but *KLF4* alone converted primed hPSCs to the ct2iLGöY state at a 2–3-fold lower efficiency when compared with the efficiency of using *OSKM* (Supplementary Fig. 4c–e and Supplementary Table 5). On a transcriptional level, HC and MDS revealed that naive ct2iLGöY K cells generated from both hESCs and hiPSCs were transcriptionally similar to naive rt2iLGöY and ct2iLGöY OSKM hPSCs and cells of the human ICM but distinct from conventional primed hESCs/hiPSCs (Fig. 5d and Supplementary Fig. 5a).

To assess homogeneity of naive ct2iLGöY OSKM, ct2iLGöY K and rt2iLGöY hiPSCs, we performed single-cell targeted transcriptional profiling for primed and naive pluripotency-related genes. Clustering by principal component analysis (PCA) showed mutually exclusive groups of human primed and naive cells. Importantly, the profiles of ct2iLGöY OSKM, ct2iLGöY K cells and rt2iLGöY cells overlapped to form a tight group (Supplementary Fig. 5b). This finding was further confirmed by Fruchterman Rheingold similarity cluster analysis (Fig. 5e). Violin plots revealed that naive cells homogeneously expressed markers previously associated with naive pluripotency, while markers for the primed state were absent or expressed at low levels (Fig. 5f). We validated the conversion using the *KLF4* mRNA transfection system. After 15 d of continuous transfections, we observed typical colonies and immunostaining patterns for ct2iLGöY K cells converted from primed hESCs/hiPSCs (Supplementary Fig. 5c,d). Collectively, we showed that primed cells can be converted into naive t2iLGöY cells with *OSKM* or *KLF4* and that the resulting cell populations were homogeneous.

Next, we investigated how *KLF4* facilitates the conversion of primed hPSCs to a naive-like state in t2iLGöY medium. FACS analysis of primed cells (no *KLF4*) or primed cells transduced with *KLF4* (+*KLF4*) followed by transition into t2iLGöY medium on day 2 revealed that by day 6, ‘no *KLF4*’ cells displayed differentiation as indicated by loss of the pluripotency markers *TRA-1-60* and *EPCAM*. Conversely, the majority of primed cells transduced with *KLF4* still expressed both markers, which indicated that *KLF4* was able to maintain the phenotype of primed cells in the t2iLGöY condition (Supplementary Fig. 6a).

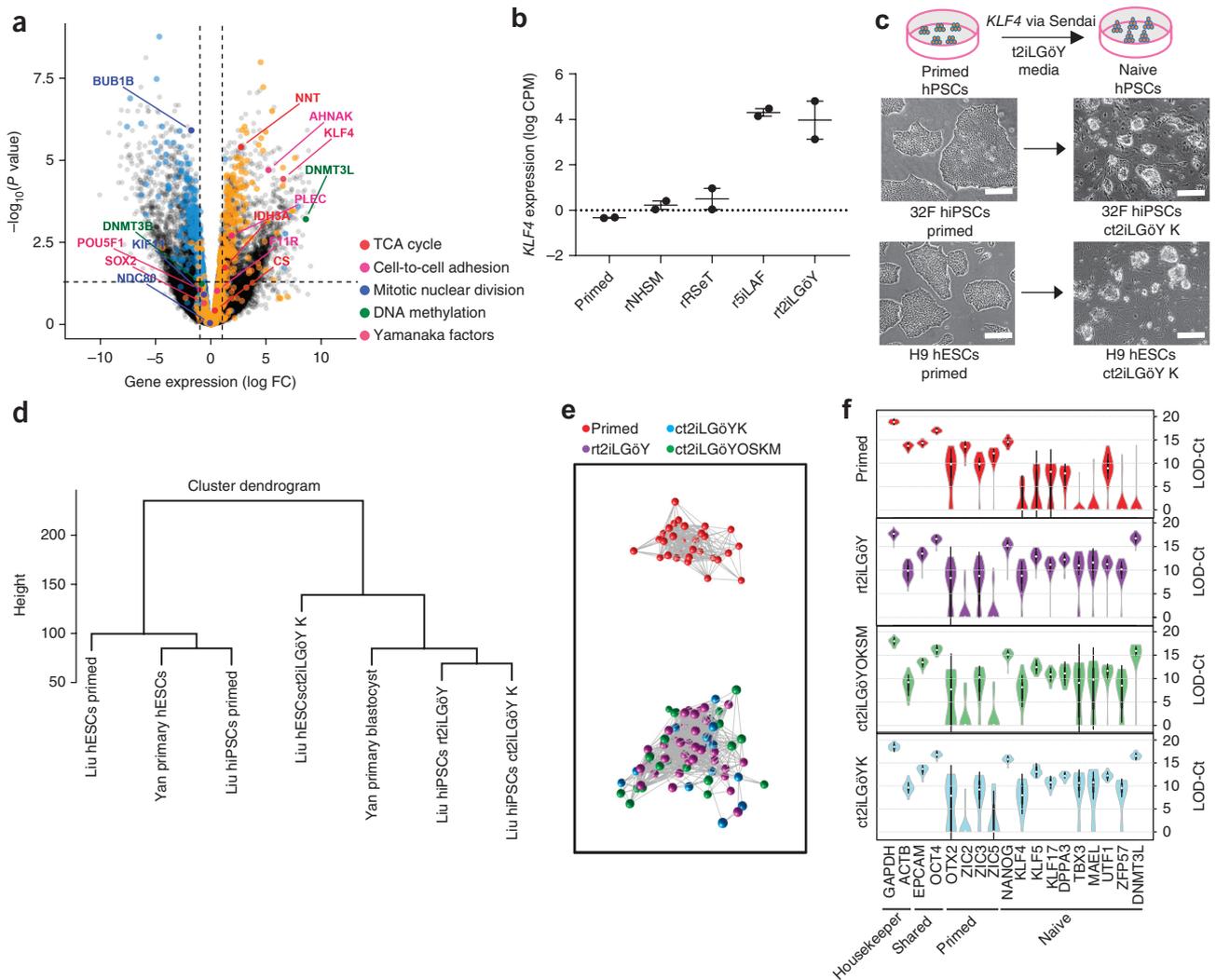


Figure 5 | Conversion of primed hPSCs to naive t2iLGöY state by *KLF4* alone. **(a)** Volcano plot depicting relationships between fold change and significance P value of expressed transcripts (RNA-seq) for primed (two iPSC cultures derived from 32F and one from 55F, $n = 3$) versus rt2iLGöY (three iPSC cultures derived from 32F, $n = 3$) and proteins (mass spectrometry; upregulated proteins indicated in orange, downregulated proteins indicated in blue) for human primed (three iPSC cultures derived from 32F, $n = 3$) versus naive rt2iLGöY (three iPSC cultures derived from 32F, $n = 3$) hiPSCs. The horizontal dashed bar represents a $-\log_{10}$ of an adjusted P value of 0.05, and vertical bars represent a \log_2 fold change of -1 and 1 . **(b)** Dot plot depicting levels of endogenous *KLF4* expression in primed, rNHSM, rRSeT, r5iLAF and rt2iLGöY hiPSCs ($n = 2$ cell donors: 32F iPSC and 38F iPSC; error bars, s.e.m.). **(c)** Schematic representation of experimental design for primed-to-naive conversion using *KLF4* alone via Sendai transduction and transition into t2iLGöY condition (top panel), and bright-field images of cultures before and after transduction. Scale bars, 250 μm (bottom panels). **(d)** Unsupervised hierarchical clustering of naive ct2iLGöY K hiPSCs/hESCs ($n = 2$ cell donors, H9 ESC and 32F iPSC) as well as primed ($n = 3$ cell donors, H9 ESC, 32F iPSC and 55F iPSC) and naive rt2iLGöY hiPSCs (two iPSC cultures derived from 32F, $n = 2$) integrated with a previously published human ICM data set²¹. **(e)** Fruchterman Rheingold similarity analysis. **(f)** Violin plots depicting the distribution of expression of a particular gene for all the single cells belonging to the specified group (white dot indicates the median, thick black line indicates the interquartile range, and the thin black line indicates the 95% confidence interval). For primed hiPSCs (32 cells), naive rt2iLGöY hiPSCs (41 cells), naive ct2iLGöY OSKM hiPSCs (19 cells) and naive ct2iLGöY K hiPSCs (18 cells), 32F background; LOD, limit of detection; Ct, cycle threshold.

Subsequently, we assessed initial transcriptional changes resulting from *KLF4* induction. We performed RNA sequencing using cells transduced with *KLF4* (and no *KLF4* controls) collected on day 2 (before transition into t2iLGöY medium) and day 6 (4 d after transition into t2iLGöY medium). Analyses revealed 242 and 178 genes differentially expressed between no *KLF4* and +*KLF4* cells on days 2 and 6, respectively. The top three gene ontology categories were extracellular matrix organization, epidermis development, and negative regulation of cell proliferation (**Supplementary Fig. 6b,c**). These categories are consistent with t2iLGöY cells growing

as dome-shaped colonies and dividing more slowly than primed hPSCs, suggesting that initial effects of *KLF4* overexpression are related to changes in cell morphology and dampening of the cell cycle. Furthermore, ACOX1, the first enzyme of the fatty acid beta-oxidation pathway, was upregulated following *KLF4* transduction on both days 2 and 6, which suggested changes in metabolism as early as day 2, before cells were changed into the t2iLGöY condition. Core naive and primed pluripotency-associated transcription factors in no *KLF4* and +*KLF4* cells on day 2 showed mild downregulation for TFs associated with the primed state

(most pronounced for OTX2) and upregulation of TFs associated with the naive state (KLF5, DPPA3, TBX3, UTF1 and DNMT3L); while, conversely, the expression of pan-pluripotency genes *OCT4*, *EPCAM* and *NANOG* were unchanged. The trend of upregulation of TFs associated with the naive state was maintained on day 6 in +*KLF4* cells; however, there was pronounced downregulation of pluripotency-associated genes *OCT4* and *NANOG* in no *KLF4* cells (Supplementary Fig. 7).

DISCUSSION

Naive mouse PSCs derived from the ICM of blastocysts can generate chimeric animals when injected into preimplantation embryos^{1,13}. Importantly, naive mouse and rat PSCs have been used to generate interspecies chimeras in proof-of-concept studies, and this has enabled the generation of a mouse pancreas in a developing rat embryo and also the generation of a rat pancreas in a developing mouse embryo^{29,30}. However, as discussed, human PSCs have commonly been cultured in primed conditions. Only recently, complex factor cocktails have been identified and used to stabilize human PSCs in naive-like states, and this resulted in the naive-permissive media evaluated in this study.

Through systematic characterization of genetically matched naive-like cells generated using different culture conditions, we have revealed cell states with distinct functional and molecular features with some more closely resembling cells from human ICM. Naive hiPSCs obtained and cultured in NHSM and RSeT conditions were mostly karyotypically stable, and they maintain correct DNA methylation patterns at primary imprinted loci and form teratomas; and, while these hiPSCs had upregulated some naive-associated genes, MDS clustered the cells near primed cells. In contrast, 5iLAF and t2iLGöY protocols generated hiPSCs closely resembling human ICM cells in terms of transcriptome and global DNA methylation levels, albeit primary imprinted DMRs were lost. Furthermore, unlike t2iLGöY hPSCs, naive 5iLAF hPSCs were karyotypically unstable in our hands.

We also found that transient expression of *OSKM* or *KLF4* alone in primed cells is sufficient for conversion to naive t2iLGöY hiPSCs. We showed that overexpression of *KLF4* rapidly triggers upregulation of naive-related genes and is required to initially maintain the pluripotent state in t2iLGöY conditions. Notably, similar conversion of EpiSCs to ESCs was achieved with expression of *Klf4* alone in mouse 2iL culture conditions^{27,28}, and knockdown of *KLF4* disrupted naive t2iL+Gö hESCs⁴. Collectively, this indicates an essential role for *KLF4* in controlling human naive gene circuitry and suggests that further studies are needed to delineate the regulatory mechanisms of human naive pluripotency.

The work presented here provides an alternate methodology to directly generate karyotypically stable naive hiPSCs that are similar to human ICM using *OSKM* to reprogram fibroblasts or *OSKM/KLF4* to convert primed hPSCs in t2iLGöY conditions. How important these characteristics are is dependent on the downstream application. For example, despite having no clear ICM signature, recent reports indicate that NHSM cells integrate into early pig embryos more efficiently than naive-like cells generated by the forced expression of *NANOG* and *KLF2* (ref. 29); this implies that trans-species chimerism may not be dependent on an ICM-like signature. The molecular and functional differences observed in these naive hPSCs suggest that pluripotency exists in a continuum. The different media studied captured cell populations

at related but distinct regions of this continuum, and this suggests fluidity in the characteristics of these cells related to media composition rather than a fully defined discrete state of human pluripotency. We envision this study will facilitate the establishment of conditions allowing generation of cells fully resembling human naive pluripotent cells of the early embryo and serve as a model system for mechanistic and clinical studies.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the high-quality cell sorting service and technical input provided by Monash Flowcore Facility and J. Hatwell-Humble for conducting the teratomas experiments. We also thank the Cytogenetics department of Monash Pathology for help with G-banding karyotype analysis. Furthermore, the authors thank the ACRF Centre for Cancer Genomic Medicine at the MHTP Medical Genomics Facility for assistance with next-generation library preparation and Illumina sequencing. This work was supported by National Health and Medical Research Council (NHMRC) project grants APP1104560 to J.M.P. and A.L.L., APP 1069830 to R.L. and a Monash University strategic grant awarded to C.M.N. X.L. was supported by a Monash International Postgraduate Research Scholarship and a Monash Graduate Scholarship. A.S.K. was supported by an NHMRC Early Career Fellowship APP1092280. J.M.P. and R.L. were supported by Silvia and Charles Viertel Senior Medical Research Fellowships.

AUTHOR CONTRIBUTIONS

X.L., C.M.N. and J.M.P. conceived the study and designed experiments. X.L. performed somatic cell reprogramming in different culture conditions and tissue culture experiments with support from C.M.N., J.C., J.F., K.M., J.M.P. and M.H.-G. S.K.N. supported the teratomas experiments. X.L. and C.M.N. performed FACS experiments, SPADE/visNE and the molecular experiments of the cells with support from A.S.K. and J.C.; A.S.K., C.H. and R.B.S. performed cell preparation for mass spectrometry experiments and analysis. F.J.R., S.M.W. and C.M.N. analyzed Fluidigm and RNA sequencing data under the guidance of J.M.P. J.P., F.J.R., E.F. and R.L. performed targeted methylation experiments and analyzed the results. N.J. and M.E.B. performed molecular characterization. H.S.C. and C.M.O'B. provided reagents and technical assistance. A.L.L. provided reagents and assisted with writing. X.L., C.M.N. and J.M.P. wrote the manuscript. All authors approved of and contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
- Chan, Y.-S. *et al.* Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* **13**, 663–675 (2013).
- Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
- Takahima, Y. *et al.* Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
- Theunissen, T.W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).
- Ware, C.B. *et al.* Derivation of naive human embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **111**, 4484–4489 (2014).
- Guo, G. *et al.* Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. *Stem Cell Reports* **6**, 437–446 (2016).

8. Huang, K., Maruyama, T. & Fan, G. The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell* **15**, 410–415 (2014).
9. Pastor, W.A. *et al.* Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* **18**, 323–329 (2016).
10. Theunissen, T.W. *et al.* Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515 (2016).
11. Sahakyan, A. *et al.* Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87–101 (2017).
12. Vallot, C. *et al.* XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell Stem Cell* **20**, 102–111 (2017).
13. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J.H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* **17**, 155–169 (2016).
14. Betschinger, J. *et al.* Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* **153**, 335–347 (2013).
15. Shakiba, N. *et al.* CD24 tracks divergent pluripotent states in mouse and human cells. *Nat. Commun.* **6**, 7329 (2015).
16. O'Brien, C.M. *et al.* New monoclonal antibodies to defined cell surface proteins on human pluripotent stem cells. *Stem Cells* **35**, 626–640 (2017).
17. Collier, A.J. *et al.* Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. *Cell Stem Cell* **20**, 874–890.e7 (2017).
18. Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* **29**, 886–891 (2011).
19. Amir, A.D. *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
20. Nefzger, C.M. *et al.* A versatile strategy for isolating a highly enriched population of intestinal stem cells. *Stem Cell Reports* **6**, 321–329 (2016).
21. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
22. Okae, H. *et al.* Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet.* **10**, e1004868 (2014).
23. von Meyenn, F. *et al.* Comparative principles of DNA methylation reprogramming during human and mouse *in vitro* primordial germ cell specification. *Dev. Cell* **39**, 104–115 (2016).
24. Bock, C. *et al.* Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).
25. Tsankov, A.M. *et al.* Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344–349 (2015).
26. Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3151–3165 (2015).
27. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069 (2009).
28. Han, D.W. *et al.* Direct reprogramming of fibroblasts into epiblast stem cells. *Nat. Cell Biol.* **13**, 66–71 (2011).
29. Wu, J. *et al.* Interspecies chimerism with mammalian pluripotent stem cells. *Cell* **168**, 473–486.e15 (2017).
30. Yamaguchi, T. *et al.* Interspecies organogenesis generates autologous functional islets. *Nature* **542**, 191–196 (2017).