TGFβ signalling is required to maintain pluripotency of human naïve pluripotent stem cells

Anna Osnato[1,2], Stephanie Brown[1,2], Christel Krueger[3], Simon Andrews[3], Amanda J. Collier[4], Shota Nakanoh[1,2,5], Mariana Quiroga Londoño[1,2], Brandon T. Wesley[1,2], Daniele Muraro[1,2,6], Sophie Brumm[7], Kathy Niakan[7,8], Ludovic Vallier[1,2,9], Daniel Ortmann[1,2,9], Peter J. Rugg-Gunn[1,4,8,9]

[1]Wellcome–MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, CB2 0AW, UK
[2]Department of Surgery, University of Cambridge, Cambridge, CB2 0SZ, UK
[3]Bioinformatics Group, The Babraham Institute, Cambridge, CB22 3AT, UK
[4]Epigenetics Programme, The Babraham Institute, Cambridge, CB22 3AT, UK
[5]Division of Embryology, National Institute for Basic Biology, Okazaki 444-8787, Japan
[6]Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK
[7]Human Embryo and Stem Cell Laboratory, The Francis Crick Institute, London, NW1 1AT, UK
[8]Centre for Trophoblast Research, University of Cambridge, Cambridge, CB2 3EG, UK
[9]These authors jointly supervised this work

Abstract

The signalling pathways that maintain primed human pluripotent stem cells (hPSCs) have been well characterised, revealing a critical role for TGFβ/Activin/Nodal signalling. In contrast, the signalling requirements of naïve human pluripotency have not been fully established. Here, we demonstrate that TGFβ signalling is required to maintain naïve hPSCs. The downstream effector proteins – SMAD2/3 – bind common sites in naïve and primed hPSCs, including shared pluripotency genes. In naïve hPSCs, SMAD2/3 additionally bind to active regulatory regions near to naïve pluripotency genes. Inhibiting TGFβ signalling in naïve hPSCs causes the downregulation of SMAD2/3-target genes and pluripotency exit. Single-cell analyses reveal that naïve and primed hPSCs follow different transcriptional trajectories after inhibition of TGFβ signalling. Primed hPSCs differentiate into neuroectoderm cells, whereas naïve hPSCs transition into trophectoderm. These results establish that there is a continuum for TGFβ pathway function in human pluripotency spanning a developmental window from naïve to primed states.

Introduction

Human pluripotent stem cells (hPSCs) are grown *in vitro* as two broadly different states termed naïve and primed (Davidson et al., 2015; Weinberger et al., 2016). The two states diverge in their embryonic identity with primed hPSCs recapitulating post-implantation epiblast, and naïve hPSCs resembling pluripotent cells of pre-implantation embryos (Rossant and Tam, 2017; Weinberger et al., 2016). This difference has profound consequences on the cell's properties, including the epigenetic state and differentiation capacity (Dong et al., 2019). Naïve hPSCs have decreased DNA methylation levels, altered distribution of histone marks, and two active X-chromosomes, and they have a higher propensity to differentiate into extraembryonic tissues (Castel et al., 2020; Cinkornpumin et al., 2020; Dong et al., 2020; Guo et al., 2021; Io et al., 2021; Linneberg-Agerholm et al., 2019; Pastor et al., 2016; Sahakyan et al., 2017; Takashima et al., 2014; Theunissen et al., 2016; Vallot et al., 2017). On the other hand, primed hPSCs represent the last stage before differentiation into the three definitive germ layers – ectoderm, mesoderm and endoderm – from which the adult organs are derived (Weinberger et al., 2016).

Importantly, these pluripotent states are established by using specific and distinct culture conditions (Taei et al., 2020). Of particular interest, primed hPSCs rely on TGFβ/Activin/Nodal signalling to maintain their self-renewal and differentiation capacity (James et al., 2005; Vallier et al., 2005). Inhibition of these pathways or knock down of their effectors – SMAD2/3 – result in the rapid differentiation towards the neuroectoderm lineage (Smith et al., 2008). Conversely, an increased activity of these signalling pathways is necessary for endoderm differentiation (D'Amour et al., 2005; Touboul et al., 2010). The mechanisms behind these apparent divergent functions remain to be fully uncovered, but the capacity of SMAD2/3 to switch partners during differentiation is likely to play a key role (Brown et al., 2011). Of note, Epiblast Stem Cells (EpiSCs) derived from post-implantation mouse embryos also rely on TGFβ/Activin/Nodal signalling (Brons et al., 2007). Furthermore, genetic studies in the mouse have shown that Nodal signalling is necessary to block neuroectoderm differentiation and to maintain the expression of pluripotency markers in the post-implantation epiblast (Camus et al., 2006; Mesnard et al., 2006). Thus, the central role of TGFβ/Activin/Nodal in primed pluripotency seems to be evolutionary conserved and is important for normal development.

In contrast, the function and evolutionary conservation of TGFβ/Activin/Nodal signalling in pre-implantation embryos is less well understood. TGFβ/Activin/Nodal signalling does not have an essential role in forming the pre-implantation epiblast in mouse (Brennan et

al., 2001; Varlet et al., 1997), whereas recent studies have suggested that the same pathway may be necessary for epiblast development in human blastocysts (Blakeley et al., 2017). The mechanistic basis for these observations are unclear. Moreover, it also remains to be established whether TGFβ signalling is required to maintain naïve hPSCs, which are the *in vitro* counterparts of pre-implantation epiblast cells. In general, naïve pluripotency is believed to be a steady state induced predominantly by blocking differentiation signals. However, the culture conditions vary between laboratories, although interestingly, most media that support naïve hPSCs contain exogenous TGFβ/Activin or a source of TGFβ provided by inactivated fibroblasts or Matrigel-coated substrates (Bayerl et al., 2021; Chan et al., 2013; Gafni et al., 2013; Guo et al., 2016; Takashima et al., 2014; Theunissen et al., 2014). Collectively, these observations suggest there could be an uncharacterised role for TGFβ/Activin/Nodal signalling specifically in the human naïve pluripotent state.

Here we address this hypothesis by first establishing that TGFβ/Activin/Nodal signalling is active in naïve hPSCs. Using genome-wide analyses, we then show that SMAD2/3 is bound near to genes that characterise the naïve pluripotent state. Furthermore, loss of function experiments demonstrate that this signalling pathway is necessary to maintain the expression of key pluripotency genes, such as *NANOG*. We then perform single-cell RNA sequencing analyses on naïve and primed hPSCs that are undergoing differentiation following the inhibition of TGFβ/Activin/Nodal signalling. In these conditions, primed hPSCs rapidly decrease pluripotency markers and activate neuroectoderm genes, whereas naïve hPSCs induce trophectoderm markers. Importantly, these analyses also suggest that SMAD2/3 directly maintains an important part of the transcriptional network characterising the naïve state. Taken together, these results show that TGFβ/Activin/Nodal signalling is necessary to maintain the pluripotent state of naïve hPSCs through directly sustaining the expression of key pluripotency genes. These new insights suggest that the function of TGFβ/Activin/Nodal signalling in human pluripotency extends to earlier stages of development than previously anticipated, thereby underlying a key species divergence that could facilitate the identification and the isolation of pluripotent states *in vitro*.

Results

TGFβ signalling pathway is active in human naïve pluripotent cells

To assess whether the key effectors of the TGFβ signalling pathway are expressed in naïve hPSCs and to evaluate the cell heterogeneity in their expression (Figure 1a; Figure supplement 1a, b), we performed single cell transcriptomic analysis (scRNA-seq) in naïve and primed hPSCs (Figure 1b, Figure supplement 1c). As expected, naïve and primed hPSCs clustered separately based on their transcriptomes. All cells expressed pan-pluripotency genes, such as *POU5F1* (also known as *OCT4*), *NANOG* and *SOX2*, however, naïve cells uniquely expressed known naïve cell markers, such as *DPPA5* and *KLF4,* and primed cells expressed *CD24* and *ZIC2* (Figure 1c, d; Figure supplement 1c). In addition, differential expression analysis confirmed the specific expression of naïve hPSCs genes, such as *KLF4*, *DPPA3* and *TFCP2L1*, and primed hPSCs factors including *DUSP6, ZIC2* and *TCF4* (Figure 1d). Importantly, we found that most TGFβ pathway effectors, such as Activin receptors (*ACVR*s) and *SMAD2-4*, are expressed at similar levels in both pluripotent cell types (Figure 1e). Interestingly, several components, including *NODAL* and *GDF3*, have higher expression levels in naïve compared to primed hPSCs (Figure 1e).

We next examined RNA-seq datasets that covered different stages of human pluripotency in stem cell lines and in embryos (Figure 1f). We first compared naïve and primed hPSCs (Collier et al., 2017) and, consistent with our scRNA-seq data, we found that most ligands, transduction proteins and receptors of the TGFβ pathway are expressed at similar levels in the two cell types (Figure 1f). Higher expression of the TGFβ ligands *NODAL* and *GDF3* and the co-receptor *TDGF1* was again detected in naïve hPSCs. Interestingly, the expression of pathway inhibitors differed, whereby *LEFTY1* and *LEFTY2* were higher in naïve hPSCs, whereas *CER1* and *FST* were higher in primed hPSCs. We then looked at gene expression changes that occur during the process of capacitation, because the transition from naïve to primed hPSCs recapitulates pre- to post-implantation epiblast cell development (Rostovskaya et al., 2019). We found that most of the effectors of the TGFβ pathway are expressed throughout the entire developmental series, and also confirmed that *NODAL* and *GDF3* are expressed at higher levels in the early stages (Figure 1f).

To examine transcriptional events directly in human embryos, we next looked at scRNA-seq data in human pre-implantation embryos from day 3 to day 7 (Petropoulos et al., 2016). Low level expression of most TGFβ pathway effectors was detected in the early inner cell mass (ICM), and their expression increased substantially in the pre-implantation epiblast

(EPI). In particular, *NODAL* and *GDF3* are highly expressed in EPI at this stage, similar to the transcriptional patterns in naïve hPSCs (Figure 1f). However, in contrast to EPI, most pathway components are undetectable in trophectoderm (TE and early TE), and are expressed at low levels in primitive endoderm (PE) (Figure 1f). These observations were extended by examining the expression of TGFβ pathway genes in a blastocyst-culture system that recapitulates EPI development from pre-implantation to early gastrulation (Xiang et al., 2020). Here, in EPI cells at 6 days post-fertilisation, *NODAL*, *GDF3* and the NODAL co-receptor *TDGF1* are highly expressed, in line with the EPI stage from the Petropoulos et al. dataset, and the high expression of these genes is sustained in all EPI cells over the following eight days of development (Figure 1f). Taken together, these results show that most ligands, transduction proteins and receptors of the TGFβ pathway are expressed at similar levels in naïve and primed hPSCs, and that this expression pattern across pluripotent states is also observed in human embryos cultured *in vitro*.

To further confirm these observations at the protein level, we examined cell-surface proteomic (Wojdyla et al., 2020) and whole-cell proteomic (Di Stefano et al., 2018) data in naïve and primed hPSCs. This revealed that most Activin/TGFβ receptors and downstream effectors of the pathways are expressed at very similar levels in the two cell types (Figure 1g, Figure supplement 1d). Finally, to directly assess TGFβ pathway activation, we performed western blot analysis and found that phospho-SMAD2 (pSMAD2), the activated form of SMAD2, is detectable in multiple embryo-derived and reprogrammed naïve hPSCs lines, and at comparable levels to primed cells (Figure 1h; Figure supplement 1e, f). The phosphorylation signal was rapidly diminished following the treatment of the cells with SB-431542 (SB), a potent and selective inhibitor that blocks TGFβ/Activin receptors ALK5, ALK4, and ALK7 (Inman et al., 2002) (Figure 1h; Figure supplement 1e, f). Taken together, these results establish that the TGFβ signalling pathway is active in naïve hPSCs. Because primed hPSCs rely on this pathway to maintain pluripotency, our findings raise the possibility that naïve hPSCs might also require TGFβ signalling to sustain their undifferentiated state.

SMAD2/3 binding is enriched at active enhancers in human naïve cells

Having established that the TGFβ signalling pathway is active in naïve hPSCs, we next profiled the genome-wide occupancy of the main downstream effectors – SMAD2/3 – using chromatin immunoprecipitation combined with genome-wide sequencing (ChIP-seq) in naïve and primed hPSCs. This analysis revealed that SMAD2/3 binding is enriched in naïve cells to a similar degree as in primed cells, as shown by independent peak calling in the two cell types (Figure

2a; Figure supplement 2a). Here, we observed regions bound by SMAD2/3 in both cell types, and also a substantial number of loci that appear to have cell type-specific binding. Importantly, canonical target genes, such as *LEFTY1/2*, *NODAL*, *NANOG* and *SMAD7*, were bound by SMAD2/3 in both cell types (Figure 2b; Figure supplement 2b), suggesting that TGFβ is active and it signals through the canonical cascade in both naïve and primed hPSCs.

In addition to the shared targets, differential binding analyses revealed over 2,000 SMAD2/3-bound sites that differed between the two cell types (Figure 2c, d; Figure supplement 2c, d). Excitingly, further examination of these differential sites revealed that in naïve hPSCs SMAD2/3 uniquely bound near to naïve-specific pluripotency genes including *DNMT3L*, *TFAP2C*, *CBFA2T2*, *KLF4* and *CDK19* (Figure 2e; Figure supplement 2d, e). Interestingly, these sites often overlapped with accessible chromatin regions and H3K27ac marks, which are signatures that are associated with active enhancers (Heintzman et al., 2009) (Figure 2e; Figure supplement 2e). In contrast, primed-specific SMAD2/3 sites were located near to genes that regulate mesendoderm differentiation, such as *TBXT*, *EOMES* and *GATA4*, or primed-state pluripotency, such as *OTX2* (Figure 2e, Figure supplement 2e). These sites correspond mostly to accessible chromatin and to regions marked by H3K4me3 and H3K27me3 signals, which typically mark the promoters of developmental genes (Azuara et al., 2006; Bernstein et al., 2006; Heintzman et al., 2009) (Figure 2e; Figure supplement 2e). These findings are supported by global analysis using ChromHMM-based chromatin state annotations (Chovanec et al., 2021), where we found that most SMAD2/3 peaks are indeed within active chromatin regions, consisting mainly of gene promoters and enhancers (Figure 2f). Interestingly, naïve-specific SMAD2/3 peaks are slightly more enriched at active enhancers compared to primed-specific peaks (30.6% vs 21.4%), and primed-specific SMAD2/3 peaks are instead more enriched at promoters (46% vs 26.5%) (Figure 2f; Figure supplement 2f).

There are widespread differences in enhancer activity between naïve and primed hPSCs (Barakat et al., 2018; Battle et al., 2019; Chovanec et al., 2021) and so to determine how changes in SMAD2/3 occupancy tracks with enhancer status we compared chromatin marks at naïve-specific SMAD2/3 sites between the two cell types. The vast majority of sites that lose SMAD2/3 occupancy in primed hPSCs also show a strong reduction in chromatin accessibility and H3K27ac/H3K4me1 signals, which suggests that SMAD2/3-bound enhancers are decommissioned in primed hPSCs (Figure 2g). Chromatin marks that denote promoters and heterochromatin regions are generally low at naïve-specific SMAD2/3 sites and are largely

unchanged in primed hPSCs, further reinforcing the connection between SMAD2/3 occupancy and active enhancers in naïve hPSCs (Figure 2h).

To obtain a more complete view of the pluripotency transcriptional network, we also overlapped SMAD2/3 peaks in naïve cells with OCT4, SOX2 and NANOG (OSN) binding (Chovanec et al., 2021). We found that OSN signals were strongly reduced at naïve-specific SMAD2/3 sites in primed hPSCs, confirming the integration of SMAD2/3 within the naïve transcription factor network (Figure 2i). Importantly, regions bound by SMAD2/3 and OSN overlapped with state-specific enhancers that are marked by open chromatin and H3K27ac, as shown for the *KLF4* and *DNMT3L* loci in naïve hPSCs, and for *OTX2* and *TBXT* in primed hPSCs (Figure 2e). Finally, to further characterise the differentially bound loci, we performed differential motif enrichment to investigate whether different binding partners might regulate SMAD2/3 binding in naïve and primed cells. Interestingly, motifs that are relatively enriched at SMAD2/3 sites in naïve compared to primed cells included NF2L1 (also known as NRF1), TFAP2A/C, KLF4 and FOXH1 (Figure 2j).

Altogether, these data suggest that SMAD2/3, the main effector of TGFβ pathway, is integrated in the naïve pluripotency network by targeting OSN-bound active enhancers that are in close proximity to key regulators of naïve pluripotency.

Inhibiting TGFβ signalling induces loss of pluripotency in human naïve cells

After establishing that the TGFβ signalling pathway could maintain directly the transcriptional network characterising human pluripotency, spanning from naïve to primed states, we next examined whether the pathway is functionally required to sustain naïve hPSCs in an undifferentiated state. We first measured the transcriptional changes that occurred in response to SB-mediated loss of pSMAD2 and inhibition of the TGFβ pathway (Figure 3a; Figure Supplement 3a, b). After only two hours of SB treatment (t2iLGö medium supplemented with SB), naïve hPSCs showed a significant reduction in the expression of the pluripotency gene *NANOG*, which is a short time frame that is consistent with *NANOG* being a direct target of SMAD2/3 signalling (Vallier et al., 2009; Xu et al., 2008) (Figure 3a, Figure supplement 3a). Other canonical downstream target genes, such as *LEFTY1/2* and *SMAD7*, were also strongly downregulated and their expression was completely abolished after 24 hours in the case of *LEFTY1/2*. Excitingly, naïve pluripotency marker genes that are bound by SMAD2/3 including *DPPA3, DPPA5*, *KLF4* and *DNMT3L* were also downregulated following SB treatment, indicating that the naïve state is disrupted in these conditions (Figure 3a, Figure supplement 3a). These results were independently validated by depleting SMAD2/3 expression using the

OPTiKD system (Bertero et al., 2016). Here, we generated stable naïve hPSCs with tetracycline (TET) inducible co-expression of shRNAs that target *SMAD2* and *SMAD3* transcripts (Figure 3b). Treating these cells with TET induced the rapid loss of *SMAD2/3* mRNA (Figure 3c), and a concomitant and significant downregulation in the expression of SMAD2/3 target genes, such as *LEFTY2*, *NODAL* and *NANOG* (Figure 3c). We also detected a significant decrease in *POU5F1* expression following SMAD2/3 knockdown and after SB treatment, suggesting that naïve hPSCs are destabilised and are exiting the pluripotent state (Figure 3a, c).

Interestingly, adding SB to naïve culture media also induced a change in cell morphology whereby naïve hPSCs lost their typical dome-shaped morphology after 3 to 5 days, and this was accompanied by the appearance of flat colonies that gradually took over the culture (Figure 3d, Figure supplement 3c). This striking phenotypic change was confirmed in a second naïve hPSCs line (Figure supplement 3d). Intriguingly, the morphology of these flat colonies resemble human trophoblast cells (Okae et al., 2018). To further investigate this, we grew naïve hPSCs for 14 days in the presence of SB and then examined the expression of trophoblast marker genes (Figure 3e). We found there was a strong upregulation in the expression of the trophectoderm marker *HAND1* and also of *TP63*, *MMP2* and *SDC1* that mark cytotrophoblast (CTB), extravillous trophoblast (ETV) and syncytiotrophoblast (STB) cell types, respectively (Figure 3e). These results were further supported by the clear reduction in NANOG protein expression following 3 to 5 days of treating naïve hPSCs with SB, in correspondence with the exit from naïve pluripotency and the appearance of the trophoblast-like colonies (Figure 3f). NANOG downregulation together with the appearance of trophoblast-like colonies was also observed in a second naïve cell line upon SB treatment (Figure supplement 3e). Importantly, the flat cell colonies also expressed typical trophoblast-associated proteins – GATA3 and HAND1 (Figure 3g, Figure supplement 3f, g).

To further characterise these cells and to investigate their ability to differentiate into trophoblast derivatives, we cultured naïve hPSCs in the presence of SB for 5 days and then transferred the cells into trophoblast stem cell (TSC) media (Dong et al., 2020; Okae et al., 2018). Although the cell population initially appeared heterogeneous, following exposure to TSC conditions the cells rapidly and uniformly acquired a homogeneous TSC-like morphology. The cells expressed TSC markers, such as GATA3 and CK19 (Figure 3h) and *CK7*, *ITGA6* and *TP63* (Figure 3i), and could be passaged and maintained in these conditions with stable growth and morphology. Naïve-derived TSCs were then induced to differentiate by switching the cells to STB and EVT media (Dong et al., 2020). This led to the downregulation

of TSC genes and the upregulation of STB and EVT markers, such as SDC1 and HLA-G, respectively (Figure 3h, i).

Taken together, these results show that blocking TGFβ signalling in naïve hPSCs rapidly destabilises the pluripotency network and allows the cells to undergo differentiation towards trophoblast-like cells, including those that can give rise to multipotent, proliferative TSCs.

Single-cell transcriptional analysis reveals a trophoblast-like population arising in response to TGFβ inhibition in human naïve cells

We next sought to investigate the processes in which TGFβ pathway inhibition drives naïve hPSCs out of their pluripotent state and towards a trophoblast phenotype. Following SB treatment, we observed that the early-stage cultures contained a heterogeneous mixture of cell morphologies that included naïve-like colonies and the flat, TSC-like colonies described above (Figure 3d, Figure supplement 3c, d). The proportion of NANOG-positive cells declined following SB treatment, with variable expression within individual colonies (Figure 3f). We also observed heterogeneous colonies that contained cells expressing the pluripotency marker OCT4 and TSC-like markers HAND1/GATA3 (Figure 3g). Because the population heterogeneity could mask important changes in cell phenotype, we used scRNA-seq to examine the effect of TGFβ inhibition over seven days of SB treatment in naïve hPSCs (Figure 4a). In addition, to better characterise the divergent developmental potential between different human pluripotent states, we compared this response to the response when primed hPSCs were treated with SB. Our aim was to investigate the trajectory of naïve hPSCs moving into a putative TSC-like population, in contrast with the neuroectodermal differentiation that is induced in primed hPSCs when TGFβ is inhibited (Vallier et al., 2009).

In both cell types, there was a clear transcriptional trajectory moving from day 0 to day 7 of SB treatment (Figure 4b, Figure supplement 4a). Importantly, there was little overlap in their trajectories (Figure 4c), confirming that the inhibition of TGFβ signalling in these two different developmental stages results in divergent differentiation processes. Louvain clustering of the combined datasets also showed separated clusters in the naïve and primed time course samples (Figure supplement 4b). Specifically, TGFβ inhibition in naïve hPSCs induced the expression of TSC-like markers, such as *HAND1*, *GATA2* and *GATA3*, whereas inhibition in primed hPSCs induced neuroectoderm markers, such as *SOX10*, *PAX6* and *LEF1* (Figure 4d, Figure supplement 4c). Interestingly, Louvain clustering of the naïve cell dataset initially follows the day 0 (Cluster A) and day 1 (Cluster B) timepoints and then resolves the

mixed population at days 3, 5 and 7 into three separate clusters (C, D and E) (Figure 4e, Figure supplement 4d). This analysis suggests that the mixed population is formed from an early differentiating population (cluster C), a transition population (cluster D), and a later-stage differentiated population (cluster E), thereby confirming a stepwise process marked by different intermediate stages.

Examining individual genes revealed the dynamics of the differentiation trajectory. Pan-pluripotency and naïve-specific genes showed a gradient in their expression patterns, starting from high expression in cluster A, diminishing levels in clusters B and C, then largely absent in clusters D and E (Figure 4f, g). In contrast, trophoblast genes become activated in clusters C, D and E, with *CDX2*, *HAND1* and *GATA3* marking early, transition and late-stage differentiating populations, respectively (Figure 4f, g). *NODAL* and *LEFTY1* are expressed predominantly in cluster A and were rapidly downregulated already in cluster B (Figure 4f, g), and other TGFβ pathway genes, such as *GDF3* and *TDGF1,* are fully downregulated when cells start transitioning towards cluster D. These results confirm the effective pathway inhibition and also that blocking TGFβ signalling allows trophectoderm differentiation.

To better characterise the Louvain clusters, we examined the top 25 genes that are differentially expressed in each cluster compared to all other clusters (Figure 4g, Figure supplement 4e, f). Differentially expressed genes that are associated with cluster A, which corresponds largely to cells at day 0, include *NANOG* and *SUSD2* (Bredenkamp et al., 2019a; Wojdyla et al., 2020) in addition to the TGFβ ligand *GDF3* and receptor *TDFG1*. Interestingly, the SMAD2/3-cofactor *FOXH1* was also identified in this category and this is consistent with our prior motif analysis of the SMAD2/3 ChIP-seq data that identified FOXH1 as a putative interactor of SMAD2/3 specifically in naïve cells (Figure 2j). Genes that are differentially expressed in cluster B are enriched for metallothioneins, such as MT1/2s, which affect cell respiration, in addition to mitochondrial genes – *SLIRP* and *MTNDL4* – and the glucose pyrophosphorylase *UGP2*, suggesting that an initial response to TGFβ inhibition could involve a metabolic switch (Mathieu and Ruohola-Baker, 2017). Cells in cluster C still express pluripotency markers, such as *POU5F1* and *DPPA5*, and have upregulated the non-coding RNAs *MEG3* and *MEG8*. Cluster D clearly marks a transition population towards TSC-like cells, with the expression of *CDX1* and *CDX2*, keratins (*KRT8*, *KRT18*), and *MARCKS*, *FABP5* and *EZR* (Cambuli et al., 2014; Ralston et al., 2010). Cluster E includes keratins (*KRT8*, *KRT18*, *KRT19*), several main regulators of trophoblast development, such as *GATA2* and *GATA3* (Ralston et al., 2010), and human specific regulators, such as *VGLL1* (Soncin et al., 2018). Lastly, because recent studies have highlighted a transcriptional overlap between

trophoblast and amnion cells (Guo et al., 2021; Io et al., 2021; Zhao et al., 2021), we examined whether genes reported to be expressed by amnion cells were upregulated in our dataset. We found that most of the amnion-associated genes examined were not detectable in any of the clusters (Figure supplement 4g). Some markers, such as *CTSV* and *TPM1*, are expressed in both amnion and trophoblast, and as expected were upregulated in cluster E (Figure supplement 4g). Although it is currently challenging to separate the transcriptional profiles of trophoblast and amnion cells, this analysis suggests that TGFβ inhibition of naïve hPSCs in these conditions does not promote the induction of reported amnion cell markers. Taken together, these results confirm that TGFβ inhibition downregulates a pluripotency program and enables trophectoderm differentiation from naïve hPSCs.

To dissect the impact of TGFβ pathway inhibition on the transcriptional changes, we overlapped cluster A and E gene expression profiles with SMAD2/3 ChIP-seq peaks. We found that a small subset of differentially expressed genes have a nearby SMAD2/3 peak (Figure 4h). Of note, many of the strongest peaks are close to differentially expressed genes, and this was especially clear for genes that are downregulated upon SB treatment (Figure 4i). Interestingly, among the downregulated genes, we found that SMAD2/3 bind within 12kb of the transcriptional start sites of TGFβ downstream effectors (*NODAL*, *LEFTY1/2*, *PMEPA1*), key genes associated with naïve pluripotency (*DNMT3L*, *DPPA4*, *AIRE*, *ID1*), genes reported to inhibit trophoblast differentiation (*NLRP7*, *TET1*) (Alici-Garipcan et al., 2020; Dawlaty et al., 2011; Koh et al., 2011; Mahadevan et al., 2014), and also near to distal enhancers for other factors, such as *KLF4* and *DEPTOR* (Figure 4h). Although less prevalent, we also found SMAD2/3 binding sites close to some genes that are transcriptionally upregulated between cluster A and E, including *EFNB2* and *FOS*, and to enhancers close to *PGF* and *MFAP5*. To further assess the significance of this association, we tested how often differentially expressed genes between clusters A and E are the closest gene to a SMAD2/3 peak. Strikingly, 21% of downregulated genes are the closest gene to a SMAD2/3 binding site, which is significantly higher than the 7% of genes in a randomly-selected group of size-matched control genes ($p<2.2x10^{16}$, Figure 4j). These results suggest that the downregulation of pluripotency-associated genes following TGFβ inhibition is functionally linked to the loss of SMAD2/3 binding.

Taken together, scRNA-seq in primed and naïve cells shows that both developmental stages rely on TGFβ signalling to maintain their undifferentiated state but, upon pathway inhibition, each cell type diverges towards different trajectories. Primed cells differentiate into neuroectoderm cells whereas, in contrast, naïve cells exit pluripotency and acquire a TSC-like

fate expressing trophoblast markers and this is triggered by the deregulation of target genes that are downstream of SMAD2/3.

TGFβ inhibition in naïve hPSCs recapitulates the transcriptome of early trophoblast specification in human embryos

Having established that naïve hPSCs respond to TGFβ inhibition by shutting down the naïve pluripotency network, thereby allowing the onset of trophoblast differentiation, we next investigated whether this differentiation process follows a developmental trajectory. To do this, we applied diffusion pseudotime to our 10X scRNA-seq data (Figure supplement 5a) and examined the pseudotime trajectory across the Louvain clusters (Figure 5a). Consistent with the prior UMAP analysis, we found that the time points (days) and the clusters progressively populate the trajectory following a similar pattern from cluster A, through B and C, towards a transition population in cluster D, and lastly the more differentiated counterpart in cluster E (Figure 5a). Overlaying the diffusion pseudotime maps with the expression of known markers reveals the initial downregulation of pluripotency genes, such as *NANOG*, was followed by a sequential upregulation of trophoblast markers, such as *CDX2*, *HAND1* and *GATA3* (Figure 5b, Figure supplement 5b). Interestingly, the transitional cell population in cluster D contains a substantial proportion of cells (~15-25%) that co-express low levels of the pluripotency gene *POU5F1* and trophoblast markers, such as *CDX2* and *HAND1* (Figure 5c). We confirmed this co-expression at the protein level using immunofluorescent microscopy (Figure supplement 5c). These results indicate that trophoblast cells arise in the population through the transition of pluripotent cells to a trophoblast fate.

To further investigate the transition from naïve pluripotency to trophoblast specification, we compared our scRNA-seq data to human embryo transcriptional datasets (Xiang et al., 2020). Correlation analysis showed that cells in clusters A, B and C are transcriptionally closest to epiblast cells, in keeping with their undifferentiated status (Figure supplement 5d). The transitional population classified as cluster D has the highest correlation with ICM and TE (Figure supplement 5d). Cells in cluster E have the highest correlation with trophoblast derivatives from the pre- and early-postimplantation embryo (Figure supplement 5d).

We next focussed our analysis on the main pluripotent cell population (cluster A), the transitioning cells (cluster D) and the differentiated cells (cluster E). We compared these clusters with the embryo cell types that showed the highest transcriptional correlations to them (Figure 5d and Figure supplement 5d). Visualising single cell transcriptomes for each cell type

on a PCA plot revealed there was a good overlap between our stem cell differentiation series and the embryo lineages (Figure 5e), further supporting a transition from EPI to the trophoblast lineage. We then used the Wilcoxon Rank Sum test to identify marker genes for each embryo lineage and examined the expression pattern of those genes in cells across clusters A, D and E. Interestingly, the two datasets have remarkably similar expression patterns, whereby the progression from clusters A to D to E closely resembles the transcriptional changes from EPI to trophoblast (Figure 5f). Among the top 20 genes per cluster (Figure supplement 5e), we found genes, such as *NANOG* and *DPPA5* for cluster A / EPI, and trophoblast markers, such as *VGLL1* and *PGF* for cluster E / trophoblast, and confirmed their expression at the single cell level over the differentiation pseudotime (Figure 5g). Taken together, these results reveal that TGFβ inhibition of naïve hPSCs causes the cells to initiate a differentiation programme from pluripotency to TE-like cells and trophoblast derivatives, activating transcriptional identities similar to the embryo counterpart

Discussion

Here we show that TGFβ/Activin/Nodal signalling is active in naïve hPSCs and that this pathway is required to maintain the cells in an undifferentiated state. These findings, therefore, establish that there is a continuum for TGFβ signalling function in pluripotency spanning a developmental window from naïve to primed states (Figure 5h).

Until now, the role of TGFβ signalling in naïve hPSCs has been unclear. Activators of this pathway are often included in naïve hPSCs culture formulations (Bayerl et al., 2021; Chan et al., 2013; Theunissen et al., 2014), suggesting that this pathway could be necessary to maintain pluripotency. Accordingly, we show here that naïve hPSCs transcribe high levels of endogenous TGFβ ligands and receptors, and the pathway is activated in standard naïve cell growth conditions as demonstrated by the phosphorylation status of SMAD2/3. These findings help to interpret previous observations from several studies. For example, when testing different culture formulations, the removal of Activin from 5iLA conditions led to an increase in the spontaneous differentiation of naïve hPSCs, and also to the reduced expression of naïve genes, including *NANOG* and *KLF4* (Theunissen et al., 2014). Furthermore, supplementing HENSM media with Activin caused naïve hPSCs to express higher levels of *KLF17*, *DNMT3L* and *DPPA3* (genes that are confirmed as SMAD2/3 targets in our study) and elevated *POU5F1* distal enhancer activity, compared to the same conditions without Activin (Bayerl et al., 2021). In addition to the effect on established naïve cell lines, Activin also enhanced the kinetics of

primed to naïve hPSCs reprogramming (Theunissen et al., 2014). At the time, the authors speculated that Activin prolongs primed hPSCs in a state that is amenable to naïve reprogramming. Based on the results from our study, we propose that TGFβ signalling is required to maintain pluripotency in cells throughout primed to naïve cell reprogramming and additionally enforce the expression of genes that promote naïve hPSCs. Thus, TGFβ/Activin/Nodal signalling helps to stabilise naïve pluripotency and the addition of Activin to naïve induction and maintenance conditions is predicted to be beneficial.

At the molecular level, our analysis showed that SMAD2/3, the DNA-binding effectors downstream of TGFβ/Activin/Nodal signalling, occupied genomic sites that were common to both naïve and primed hPSCs, in addition to a large set of cell type-specific sites. Shared target genes included core pluripotency factors, such as *NANOG*, in addition to factors that are canonical targets, such as *LEFTY1/2* and *SMAD7*. Disrupting TGFβ signalling in naïve and primed hPSCs caused the rapid downregulation of these common target genes, indicating the presence of shared gene regulatory networks between the two pluripotent states. We additionally identified a large set of genes that were targeted by SMAD2/3 in naïve hPSCs but not in primed hPSCs. This set of genes included *KLF4*, *TFAP2C* and *DNMT3L*, which are important regulators of naïve pluripotency (Bayerl et al., 2021; Pastor et al., 2018), and we demonstrated that their expression levels were also sensitive to TGFβ pathway inhibition. These findings indicate that TGFβ/Activin/Nodal signalling functions in naïve hPSCs to reinforce the expression of key genes that promote naïve pluripotency, rather than to repress differentiation-promoting factors. Previous studies suggest that TGFβ/Activin/Nodal signalling may regulate NANOG expression in human embryos (Blakeley et al., 2017). It will be important to determine in the future whether the signalling requirements we uncover in naïve hPSCs could also be operating in pluripotent cells of human embryos. If so, then existing naïve hPSCs may serve as a useful cell model in which to investigate the mechanisms of signalling pathways that are relevant for early human development, alternatively, if this shows distinctions it may point to ways in which current *in vitro* conditions may need to be further refined to more closely recapitulate the pre-implantation embryonic epiblast in the embryo. Importantly, genetic studies in the mouse have established a key function for Nodal-SMAD2/3 signalling in maintaining the pluripotent state of post-implantation epiblast and in the formation of the primitive streak during gastrulation (Brennan et al., 2001; Varlet et al., 1997). Concerning pre-implantation stages, TGFβ/Activin/Nodal signalling appears to play a role in the regionalisation of the extraembryonic endoderm. However, a function in the early epiblast

remains elusive, thereby suggesting the existence of species divergence regarding TGFβ/Activin/Nodal signalling function during early development.

Our experiments also uncovered a widespread relocalisation in the genomic sites that are occupied by SMAD2/3. By integrating our datasets with chromatin and transcription factor profiles, we found that SMAD2/3 binding was enriched at active enhancers in naïve cells, yet predominantly at promoters in primed cells. This redistribution mirrors changes in OCT4, SOX2 and NANOG occupancy, whereby sites bound by SMAD2/3 only in naïve hPSCs are also preferentially occupied by OSN in naïve compared to primed cells. These findings predict that SMAD2/3 and OSN integrate signalling and transcription factor inputs in naïve pluripotency, similar to the functional interaction between SMAD2/3 and NANOG in primed hPSCs (Brown et al., 2011; Xu et al., 2008). Together, these results establish that TGFβ signalling is a core feature that is closely integrated within the transcriptional network of naïve hPSCs.

Finally, our single-cell analysis revealed that naïve and primed hPSCs depart along different trajectories following TGFβ inhibition. Primed hPSCs differentiated rapidly into neuroectoderm following TGFβ inhibition, which is consistent with previous studies (Smith et al., 2008). In contrast, naïve hPSCs upregulated trophoblast-associated genes after several days of TGFβ inhibition. The divergent routes taken by naïve and primed hPSCs could be due to their different developmental states and differentiation potential. In keeping with their preimplantation epiblast identity, naïve hPSCs can differentiate efficiently into trophoblast and hypoblast, and are required to transition through a process of capacitation to acquire the competency to respond directly to signals that promote postimplantation germ layer induction. In contrast, primed hPSCs are more similar to early postimplantation epiblast, and therefore do not efficiently make trophoblast and hypoblast. Additionally, the presence of different factors and inhibitors in the naïve and primed hPSC culture media could also affect their responses to TGFβ inhibition. Importantly, not all of the cells in the inhibitor-treated naïve cultures differentiated uniformly over the first few days. Thus, we speculate that the presence of small molecules inhibiting MEK, GSK3β, and PKC with the addition of LIF can attenuate the effect of TGFβ inhibition (Guo et al., 2021). Notably, a TGFβ inhibitor is a common component of human TSC medium (Okae et al., 2018), which suggests that TGFβ signalling may act to limit trophoblast self-renewal or proliferation. TGFβ inhibitors are also a component in the conditions that can convert naïve hPSCs to TSCs (Castel et al., 2020; Cinkornpumin et al., 2020; Dong et al., 2020; Liu et al., 2020). Here, the inhibitor might be functioning in two ways: to induce the exit from naïve pluripotency, and to promote trophoblast cell growth.

Unexpectedly, our single cell analysis revealed that following TGFβ inhibition, naïve cells acquire a transcriptional identity closest to ICM and early TE, marked, for example, by transient *CDX2* expression, and then the cells undergo further differentiation into trophoblast cell types. *CDX1* and *CDX2* are expressed transiently in primate trophoblast development including the pre-implantation TE in human blastocysts (Niakan and Eggan, 2013), but are not expressed in embryo-derived TSCs or naïve hPSC-derived TSCs, which are more similar to post-implantation trophoblast (Castel et al., 2020; Dong et al., 2020; Okae et al., 2018). By capturing pre-implantation TE-like cells, our scRNA-seq data could, therefore, shed light on the transcriptional changes that occur during the early stages of human trophoblast specification.

To conclude, our results establish a central role for TGFβ/Activin/Nodal signalling in protecting human pluripotent stem cells against differentiation. This knowledge will be useful to establish culture conditions allowing the derivation and production *in vitro* of different cell types constituting the human embryo. In addition, modulation of TGFβ could play a key role in the early human embryo and could be useful for improving culture conditions used to grow human embryos *in vitro*.

Materials and Methods

Cell culture

Transgene-reset WA09/H9 NK2, embryo-derived HNES1 and chemically-reset cR-H9 naïve hPSCs (Guo et al., 2017, 2016; Takashima et al., 2014) were kindly provided by Dr. Austin Smith with permission from WiCell and the UK Stem Cell Bank Steering Committee. Cells were maintained in t2iLGö (Takashima et al., 2014) or in PGXL (Bredenkamp et al., 2019b; Rostovskaya et al., 2019) in hypoxia (5% O_2) at 37˚C. The N2B27 base medium contained a 1:1 mixture of DMEM/F12 and Neurobasal, 0.5X N-2 supplement, 0.5X B-27 supplement, 2mM L-Glutamine, 0.1mM β-mercaptoethanol (all from ThermoFisher Scientific), 0.5X Penicillin/Streptomycin. For t2iLGö, the base medium was supplemented with 1μM PD0325901, 1μM CHIR99021, 20ng/ml human LIF (all from WT-MRC Cambridge Stem Cell Institute) and 2μM Gö6983 (Tocris). For PXGL, N2B27 medium was supplemented with 1μM PD0325901, 2μM XAV939, 2μM Gö6983 and 10 ng/ml human LIF. Naïve hPSCs were maintained on a layer of irradiated mouse fibroblasts that were seeded at a density of two million cells per 6-well plate. All experiments have been performed on naïve hPSCs that were grown in the absence of mouse fibroblasts for at least two passages using Growth Factor Reduced Matrigel-coated plates (Corning). For TGFβ inhibition experiments, 20μM SB431542 (Tocris) was added to the medium for the specified length of the experiment.

Conventional (primed) WA09/H9 (Thomson et al., 1998) were maintained in E8 medium as previously described (Chen et al., 2011) in DMEM/F12, 0.05% Sodium Bicarbonate, 2X Insulin-Transferrin-Selenium solution (all from ThermoFisher Scientific), 64μg/ml L-ascorbic acid 2-phosphate (LAA) (Sigma), 1X Penicillin/Streptomycin (WT-MRC Cambridge Stem Cell Institute), 25ng/ml FGF2 (Hyvönen Group, Dept of Biochemistry) and 2ng/ml TGFβ (BioTechne) on 10μg/ml of Vitronectin XF-coated plates (StemCell Technologies) at 37˚C. For TGFβ inhibition experiments, 10μM SB431542 (Tocris) was added to the media for the specified length of the experiment.

TSCs, ETV and STB cells were generated as previously described (Dong et al., 2020) with some modifications as follows. WA09/H9 NK2 naive hPSCs were treated with 10μM SB431542 (Tocris) in t2iLGö media for 5 days. Cells were dissociated with TrypLE (ThermoFisher Scientific) and single cells were seeded on Collagen IV-coated plates (5μg/ml; Sigma) in TSC media (Okae et al., 2018) comprising of DMEM/F12 supplemented with 0.1mM β-mercaptoethanol, 0.2% FBS, 0.5% Penicillin/Streptomycin, 0.3% BSA, 1% ITS-X (all from ThermoFisher Scientific), 1.5μg/ml L-ascorbic acid (Sigma), 50ng/ml EGF

(Peprotech), 2μM CHIR99021 (WT-MRC Cambridge Stem Cell Institute), 0.5μM A83-01 (Tocris), 1μM SB431542 (Tocris), 0.8mM VPA (Sigma), and 5μM Y-27632 (Cell Guidance Systems) in 5% CO_2. The media was changed every two days and cells were passaged with TrypLE when ~80% confluent. To induce EVT differentiation, dissociated naive-derived TSCs were seeded onto plates pre-coated with 1μg/ml of Collagen IV (Sigma) in EVT basal media comprising DMEM/F12 with 0.1mM β-mercaptoethanol, 0.5% Penicillin/Streptomycin, 0.3% BSA, 1% ITS-X (all ThermoFisher Scientific), 7.5μM A83-01 (Tocris), 2.5μM Y27632 (Cell Guidance Systems) and supplemented with 4% KSR (ThermoFisher Scientific) and 100ng/ml NRG1 (Cell Signalling). Matrigel (Corning) was added at 2% final concentration shortly after resuspending the cells in the media. On day 3, the media was replaced with EVT basal media supplemented with 4% KSR (ThermoFisher Scientific), and Matrigel (Corning) was added at 0.5% final concentration. On day 6, the media were replaced with EVT basal medium, plus 0.5% Matrigel (Corning). EVTs were cultured for two more days and then collected for analysis. To induce STB differentiation, dissociated TSCs were seeded in STB media comprising DMEM/F12 supplemented with 0.1mM β-mercaptoethanol, 0.5% Penicillin/Streptomycin, 0.3% BSA, 1% ITS-X (all ThermoFisher Scientific), 2.5μM Y-27632 (Cell Guidance Systems), 50ng/ml EGF (Peprotech), 2μM Forskolin (R&D) and 4% KSR (ThermFisher Scientific) in ultra-low attachment plates to form cell aggregates in suspension. Fresh media was added on day 3, and samples were collected for analysis on day 6.

Authentication of hPSCs was achieved by confirming the expression of pluripotency genes and protein markers (NANOG and OCT4). Cells were routinely verified as mycoplasma-free using broth and PCR-based assays. The cell lines are not on the list of commonly misidentified cell lines (International Cell Line Authentication Committee).

Western Blotting

For whole cell lysates, cells were washed once in D-PBS and resuspended in ice cold RIPA buffer (150mM NaCl, 50mM Tris, pH8.0, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate) containing protease and phosphatase inhibitors for 10min. Protein concentration was quantified by a BCA assay (Pierce) following the manufacturer's instructions using a standard curve generated from BSA and read at 600nm on an EnVision 2104 plate reader. Samples were prepared by adding 4x NuPAGE LDS sample buffer (ThermoFisher Scientific) plus 1% β-mercaptoethanol and heated at 95˚C for 5min. 5-10μg of protein per sample was run on a 4%–12% NuPAGE Bis-Tris Gel (ThermoFisher Scientific) and then transferred to PVDF membrane by liquid transfer using NuPAGE Transfer buffer

(ThermoFisher Scientific). Membranes were blocked for 1hr at RT in PBS 0.05% Tween-20 (PBST) supplemented with 4% non-fat dried milk and incubated overnight at 4˚C with primary antibodies diluted in the same blocking buffer, or 5% BSA in case of phosphor-proteins. After three washes in PBST, membranes were incubated for 1h at RT with horseradish peroxidase (HRP)-conjugated secondary antibodies diluted in blocking buffer, then washed a further three times before being incubated with Pierce ECL2 Western Blotting Substrate (ThermoFisher Scientific) and exposed to X-Ray Film. Membranes were probed with antibodies in Table 1. Relative quantification was performed using Fiji (ImageJ). Western blots were performed in three different lines, with the NK2 line in biological duplicate (Figure 1 and Supplementary Figure 1 and 3).

RNA extraction and quantitative reverse transcription PCR (RT-qPCR)

Total RNA was extracted with the GeneElute Total RNA kit (Sigma). The on-column DNase digestion step was performed (Sigma) to remove any genomic DNA contamination. 500ng of total RNA was used to synthesize cDNA with SuperScript II (ThermoFisher Scientific) using Random primers (Promega) following manufacturer's instructions. cDNA was diluted 30-fold and 2.5µl was used to perform Quantitative PCR using Kapa SYBR fast Low-Rox (Sigma) in a final reaction volume of 7.5µl on a QuantStudio 5 384 PCR machine (ThermoFisher Scientific). Samples were run in technical duplicate as two wells in the same qPCR plate and results were analysed using *PBGD*/*RPLP0* as housekeeping genes. All experiments were run in biological triplicate unless specified in the figure legends. Biological replicates were defined as separate experiments using the same line from three different passages performed at different times. All primer pairs were validated to ensure only one product was amplified and with a PCR efficiency of 100% (±10%). Primer sequences used are displayed in Table 2.

SMAD2/3 iKD line and reprogramming

Validated short hairpin RNA (shRNA) sequences against SMAD2 (CCGGCAAGTACTCCTTGCTGGATTGCTCGAGCAATCCAGCAAGGAGTACTTGTTTTTG) and SMAD3 (CCGGGCCTCAGTGACAGCGCTATTTCTCGAGAAATAGCGCTGTCACTGAGGCTTTTTG) were obtained from Sigma. Construction and transfection of the sOPTiKD plasmid as well as cloning were carried out as described in (Bertero et al., 2018). GeneJuice Transfection Reagent (Sigma) was used for transfection.

Primed SMAD2/3 inducible knockdown hPSCs were reprogrammed to a naïve state in 5i/LA conditions (Theunissen et al., 2014). Primed hPSCs were dissociated into single cells with Accutase and 1.2 million cells per 10cm tissue culture dish were plated in primed hPSCs media with 10µM Y-27632 (Cell Guidance Systems) onto MEF seeded at a density of 4 million cells per 10cm dish. The following day, media was changed to 5i/LA comprising of a 1:1 mixture of DMEM/F12 and Neurobasal, 1X N-2 supplement, 1X B-27 supplement, 1% nonessential amino acids, 2mM GlutaMAX, 50U/ml and 50µg/ml penicillin-streptomycin (all from ThermoFisher Scientific), 0.1mM β-mercaptoethanol (Millipore), 50µg/ml bovine serum albumin (ThermoFisher Scientific), 0.5% Knockout Serum Replacement (ThermoFisher Scientific), 20ng/ml recombinant human LIF, 20ng/ml ActivinA, 1µM PD0325901 (all from WT-MRC Cambridge Stem Cell Institute), 1µM IM-12, 1µM WH-4-023, 0.5µM SB590885 and 10µM Y-27632 (all from Cell Guidance Systems). Cells were passaged with Accutase on days 5 and 10. Knockdown was induced by adding 1µg/ml Tetracycline (Sigma) dissolved in Embryo Transfer Water (Sigma) to the media.

Immunofluorescence

Cells were grown on glass coverslips coated with either Matrigel or Vitronectin XF and fixed with 4% PFA for 10min at RT, rinsed twice with PBS, and permeabilised for 20min at RT using PBS/0.25% Triton X-100 (Sigma). Cells were blocked for 30min at RT with blocking solution (PBS-0.25% Triton X-100 plus BSA 1%). Primary and secondary antibodies (listed in Table 1) were diluted in blocking solution and incubated for 1h at 37˚C. Cells were washed twice with blocking solution after each antibody staining, and stained with DAPI for 5min at RT (0.1µg/ml DAPI in PBS-0.1% Triton). Finally, coverslips were mounted on slides using ProLong Gold antifade reagent (ThermoFisher Scientific) and imaged using an LSM 700 confocal microscope (Zeiss). To image STB cells, cell aggregates were collected by gentle centrifugation (100 x g for 30sec) and fixed in 4% PFA for 20min. Cells were rinsed twice with PBS and resuspended in 100µl of PBS and dried overnight on plus-charged slides (SuperFrost Plus™ Adhesion slides, Fisher Scientific). The area containing the dried cells was circled with a PAP pen and the cells were permeabilised for 5min at RT with 100µl of 0.1% Triton-X100 in PBS, and then blocked for 1h at RT with 100µl of 0.1% Triton-X100 plus 0.5% BSA. Primary antibodies (listed in Table 1) were diluted in blocking solution and incubated overnight at 4˚C. Cells were washed three times with blocking solution for 5min, and stained with secondary antibodies for 1h at RT. Cells were then washed for 15min with PBS, followed by a second PBS wash supplemented with DAPI (0.1µg/ml) and a third with PBS. Finally,

coverslips were mounted using ProLong Gold antifade reagent (ThermoFisher Scientific) and imaged using an LSM 700 confocal microscope (Zeiss). Images processed using the software Fiji (ImageJ). At least four different fields from each experiment were imaged and representative ones are shown in the figures.

Chromatin immunoprecipitation (ChIP) sequencing

Chromatin immunoprecipitation (ChIP) was performed as previously described (Brown et al., 2011), using HEPES buffer containing 1% formaldehyde at room temperature, 10mM Dimethyl 3,3′-dithiopropionimidate dihydrochloride (DTBP, Sigma) and 2.5mM 3,3′-Dithiodipropionic acid di(N-hydroxysuccinimide ester) (DSP, Sigma) for the crosslinking step. Experiments were performed on biological duplicates, carried out at different times with cells from two different passages. 10µg of SMAD2/3 antibody (Table 1) was used per ChIP, and samples were purified using the iPure v2 bead kit (Diagenode). Libraries were constructed using the MicroPlex Library Preparation Kit v2 (Diagenode) following the manufacturer's instructions, 10ng of input and all of the ChIP DNA was used as the starting material. Libraries were quantified using KAPA Library Quantification Kit (Roche) following the manufacturer's instructions and by BioAnalyser. Sequencing was performed at the Babraham Institute's Next-Generation Sequencing Facility. Equimolar amounts of each library were pooled, and eight samples were multiplexed on one lane of a NextSeq500 HighOutput 75bp Single End run.

Data processing

Reads were quality and adapter trimmed using Trim Galore! (version 0.5.0_dev, Cutadapt version 1.15), and aligned to GRCh38 using Bowtie 2 (version 2.3.2).

Data analysis

All analyses were performed using SeqMonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/, version 1.46.0) or R (https://www.R-project.org/, version 4.0.2). For quantitation, read lengths were extended to 300 bp and regions of coverage outliers were excluded. SMAD2/3 peaks were called using a SeqMonk implementation of MACS (Zhang, et. al, 2008) with parameters p<10E-6, sonicated fragment size = 300. Peaks were called individually for both replicates and the overlap of peaks used for annotation. Control regions were randomly selected from 700 bp tiles not overlapping excluded regions.

Differential binding analysis was performed using the R package Diffbind, and analysis of motifs that are relatively enriched in naïve compared to primed was performed using the MEME suit tool AME.

Single cell RNA-seq (10X Chromium Single Cell)

H9 NK2 naïve and H9 primed hESCs were grown in presence of 20µM (naïve) and 10µM (primed) of SB431542 for 7 days. Cells during the time-course were collected at day 0 (control, no SB) and at day 1, 3, 5 and 7 of treatment and dissociated with Accutase (ThermoFisher Scientific) for 5min at 37˚C in hypoxia (5% O_2), and resuspended until single cell suspension was obtained. Accutase was blocked by adding PBS/BSA 0.5% with the respective media, and after a wash pellets were resuspended at a concentration of ~1,000 cells/µl in the respective media. 3,000 cells/sample were loaded on a Chromium Chip B Single Cell following the manufacturer's instruction to generate Gel Beads-in-emulsion (GEMs) using a Reagent kit v3. Final Chromium Single Cell 3' Gene Expression library was generated using standard Illumina paired-end constructs with P5 and P7 primers.

Data analysis

Cell Ranger pipeline (version 3.0.2) was used to align reads to GRCh38 assembly and generate feature-barcode matrices for further gene expression analyses. Quality control, normalisation, dimensionality reduction analyses and all downstream analyses were carried out using the python-based library Scanpy (Wolf et al., 2018). Genes with read counts > 0 in at least 3 cells and cells expressing at least 200 genes were maintained for downstream analysis. Low quality cells were removed based on the percentage of unique molecular identifiers (UMIs) mapping to the mitochondrial genome and the number of genes detected. Logarithmic normalisation was performed, highly variable genes were selected, the total number of UMIs per cell was regressed out from log-normalised data and the regressed expression values were scaled. The dimensionality reduction was performed using Principal Component Analysis (PCA) and the neighborhood graph of cells was calculated using the PCA representation of the scaled data matrix. Clustering was performed on scaled data using the Louvain method. This graph was embedded in two dimensions using Uniform Manifold Approximation and Projection (UMAP).

Transcriptional similarity was also quantified at origin and region resolution by estimating the connectivity of data manifold partitions within the partition-based graph abstraction (PAGA) framework. Cluster markers and differentially expressed genes were

identified by applying the Wilcoxon-Rank-Sum test. In order to visualise the gradual variation in the transcriptional profile following the differentiation induced by the SB treatment, cells were represented as a pseudo-spatial dimension using the diffusion pseudotime method.

FPKM values for the ICM, EPI and TE/CTB single cell RNA-Seq datasets from (Xiang et al., 2020) (GSE136447) were extracted and log2 transformed. Similarly, log2 reads per 10k values from the first 200 cells from the A, D and E Louvain clusters were also prepared. For PCA analysis the data was filtered to retain only genes which were expressed in more than 10% of cells in both datasets. PCA was used to separate the cells in the filtered Xiang et al. data, retaining the first and second principal components. The rotations from this analysis were then applied to the Louvain cluster data to project it into the same space (Figure 5e).

For overall correlation (Figure 5d and Supplementary Figure 5d) the mean log2 FPKM for each condition from the Xiang et al. data was correlated with the summed, log2 FPM values from the Louvain clusters using Pearson's correlation. Only genes with log2 FPKM > 0.2 in any Xiang et al. dataset and raw counts > 2 in any Louvain cluster were used for the calculation.

For the single cell heatmaps (Figure 5f) a Wilcoxon Rank Sum test was used to identify marker genes which were significantly (fdr < 0.05, Benjamini-Hochberg correction) enriched in one Xiang et al. scRNA-seq condition relative to the others. We then plotted a heatmap of the expression patterns of the marker genes (columns) in each cell (rows) for both the Xiang et al. and Louvain cluster data, with the cells being ordered by the group to which they belonged. Measures were per-gene z-score normalised log2 FPM.

Data availability

Sequence data that support the findings of this study have been deposited in ArrayExpress with the accession numbers E-MTAB-10017 (ChIP-seq) and E-MTAB-10018 (scRNA-seq). Source data files have been provided for Figure 3.

Acknowledgements

We thank Kristina Tabbada and Clare Murnane of the Babraham Institute Next Generation Sequencing Facility, and Felix Krueger from Babraham Bioinformatics for sequencing QC and mapping. We also thank Steven Leonard of the Wellcome Sanger Institute for pre-processing single cell RNA-seq data. We are very grateful to Vicente Perez-Garcia (Babraham Institute and the Centre for Trophoblast Research) for providing advice and reagents for characterising the naïve-derived human trophoblast cells. Work in our laboratories is supported by grants from the BBSRC (BBS/E/B/000C0421, BBS/E/B/000C0422) and the MRC (MR/T011769/1). A.J.C. was supported by an MRC DTG Studentship (MR/J003808/1). This work was also supported by the European Research Council Grant New-Chol (L.V., A.O.), the Cambridge Hospitals National Institute for Health Research Biomedical Research Center (L.V., S.B.), the EU H2020 INTENS grant (D.O.), Gates Cambridge PhD studentship (B.T.W.), JSPS Overseas Research Fellowship (201860446) and a Grant-in-Aid (16J08005) (S.N.), and core support grant from the Wellcome and Medical Research Council to the Wellcome – Medical Research Council Cambridge Stem Cell Institute.

Competing interests

The authors declare no competing interests.

Figure 1

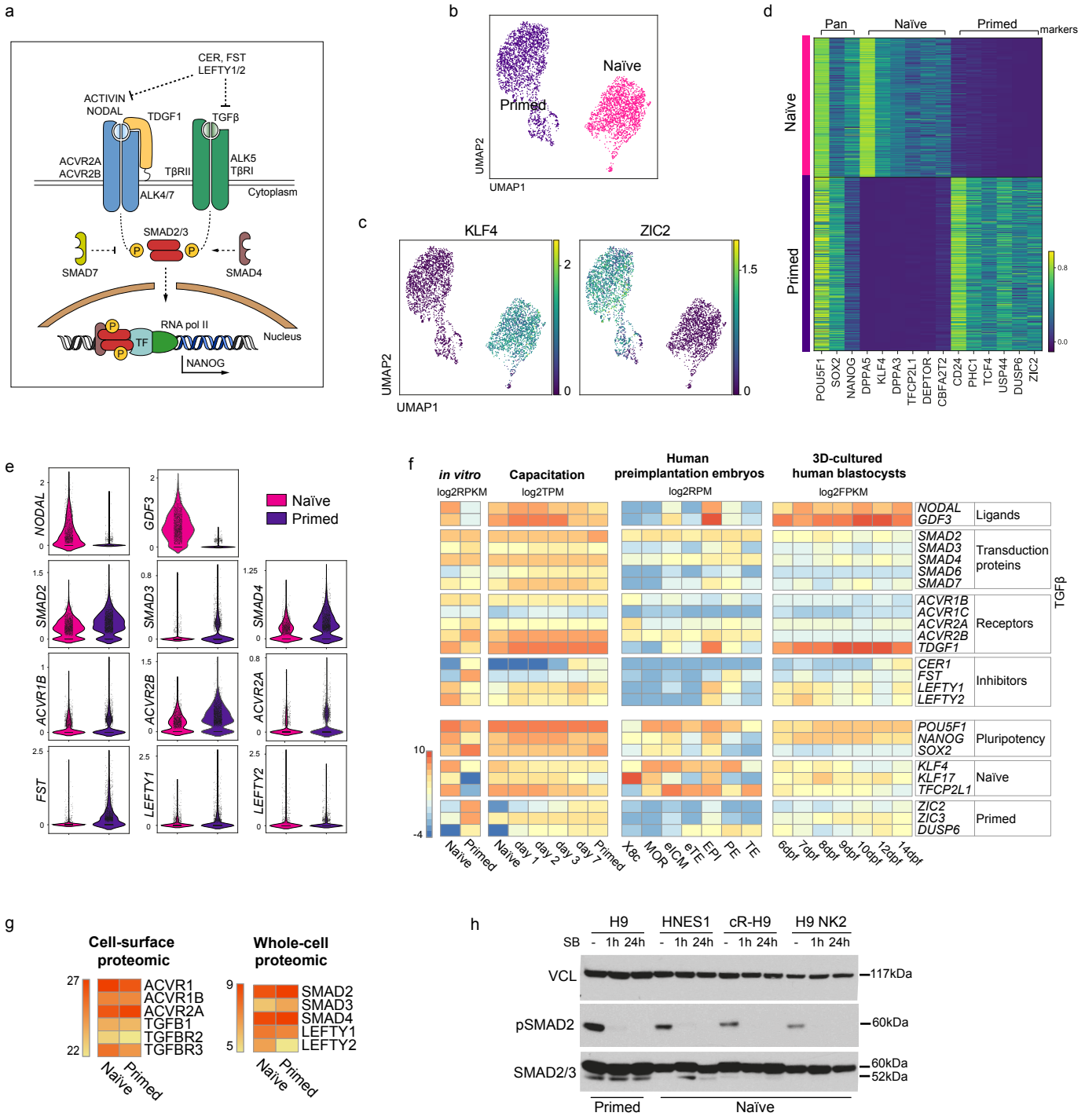


Figure 1. TGFβ signalling pathway is active in human naïve pluripotent stem cells

a) Overview of the TGFβ signalling pathway. Extracellular ligands ACTIVIN and NODAL bind to type I (ACVR2A/2B) and type II transmembrane receptors (ALK4/7), and TGFβ binds to TβRI and TβRII/ALK5. NODAL requires the additional transmembrane co-receptor TDGF1 (CRIPTO1). The activated receptor complex phosphorylates the linker region of SMAD2 and SMAD3, which enter the nucleus in complex with SMAD4. They act as transcriptional regulators and induce or repress the transcription of their target loci by recruiting other transcription factors (TF) and epigenetic modifiers. Several negative regulators of the signalling pathway are also shown: LEFTY1/2 block the signalling pathway by binding to the receptors; Cerberus (CER) and Follistatin (FST) block the ligands; SMAD7 inhibits the SMAD2/3 complex.

b) 10X RNA-seq data of naïve and primed hPSCs represented on a UMAP plot.

c) UMAP visualisation of naïve and primed hPSCs reporting the relative expression of respective pluripotent state markers, *KLF4* and *ZIC2*.

d) Heatmap reporting the expression values of selected naïve and primed marker genes divided in pan-pluripotency markers, and naïve- and primed-specific markers within the top 250 differentially expressed genes.

e) Violin plots of the 10X RNA-seq data comparing the transcript expression of TGFβ effectors in naïve and primed hPSCs.

f) Heatmap summarising the transcript expression of TGFβ effectors and pluripotency genes. RNA-seq datasets shown are: *in vitro*-cultured naïve and primed hPSCs (Collier et al., 2017), hPSCs undergoing naïve to primed state capacitation (Rostovskaya et al., 2019), human pre-implantation embryos (Petropoulos et al., 2016), and epiblast cells within a 3D human blastocyst culture system (Xiang et al., 2020). X8c: 8-cell stage; MOR: morula; eICM: early-ICM; eTE: early-trophectoderm; EPI: epiblast; PE: primitive endoderm; TE: trophectoderm. Dpf: days post-fertilisation.

g) Heatmaps summarising protein abundance levels determined by cell-surface proteomics (Wojdyla et al., 2020) and whole cell proteomics (Di Stefano et al., 2018) for TGFβ effectors in naïve and primed hPSCs.

h) Western blot analysis of TGFβ signalling pathway activation in H9 primed hPSCs (cultured in E8 medium) and in three naïve hPSC lines cultured in t2iLGö medium: embryo-derived HNES1, chemically-reset cR-H9, and transgene-reset H9 NK2. Blots show SMAD2 phosphorylation signal (pSMAD2-Ser465/Ser467) and total SMAD2/3 levels in normal

conditions (-), and following 1h and 24h of SB431542 supplementation to their culture media. Vinculin (VCL) used as a loading control.

Figure 2

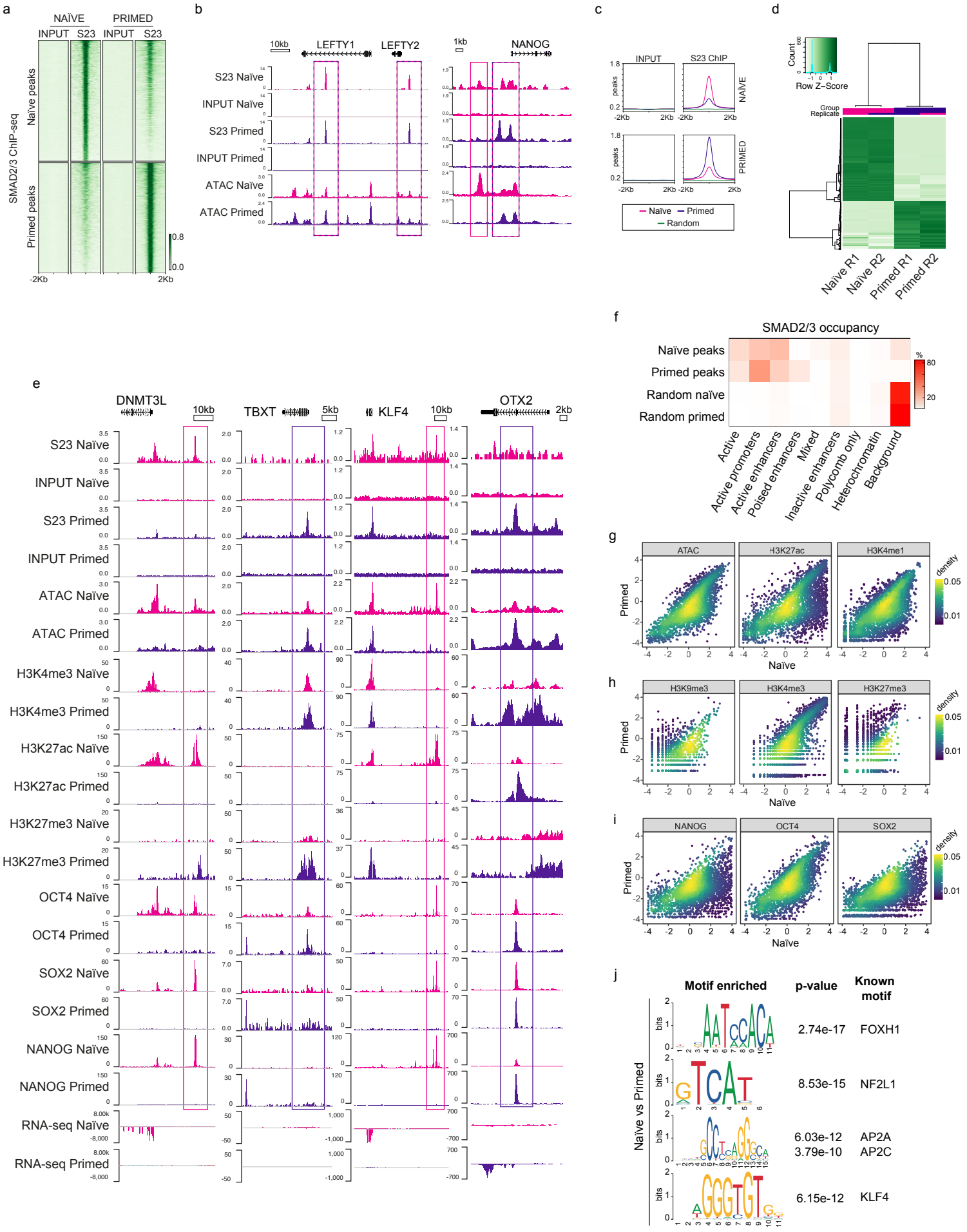


Figure 2. SMAD2/3 binds to chromatin at common and pluripotent state-specific sites

a) Heatmap displaying normalised SMAD2/3 (S23) ChIP-seq reads ±2kb from the centre of SMAD2/3-bound peaks that were independently defined in naïve (H9 NK2 line cultured in t2iLGö medium) and primed (H9 line cultured in E8 medium) hPSCs; two biological replicates per cell line. Top panel shows the regions identified as SMAD2/3-bound peaks in naïve cells; lower panel shows SMAD2/3-bound peaks in primed cells.

b) Genome browser tracks reporting SMAD2/3 (S23) binding (this study) and chromatin accessibility (ATAC-seq; Pastor et al., 2018) at the *LEFTY1/2* and *NANOG* loci in naïve and primed hPSCs. Input tracks are shown as controls.

c) Normalised average meta-plots of SMAD2/3 (S23) ChIP signal ±2kb from the centre of the peaks in naïve and primed hPSCs, compared to a randomly-selected subset of regions.

d) Heatmap displaying regions that are differentially bound by SMAD2/3 in naïve and primed hPSCs in two biological replicates (R1 and R2).

e) Genome browser tracks reporting expression (RNA-seq), chromatin accessibility (ATAC-seq), and ChIP-seq datasets of SMAD2/3 (S23), histone marks for enhancers (H3K27ac) and promoters (H3K4me3, H3K27me3), and transcription factors (OCT4, SOX2, NANOG) at the *DNMT3L*, *TBXT*, *KLF4*, *OTX2* loci. Input tracks are shown as controls. The following data sets are shown: ATAC-seq (Pastor et al., 2018); H3K4me3 (Theunissen et al., 2014); H3K4me1 (Chovanec et al., 2021; Gifford et al., 2013); H3K27me3 (Theunissen et al., 2014); H3K27ac (Ji et al., 2016); OCT4 (Ji et al., 2016); SOX2 (Chovanec et al., 2021); NANOG (Chovanec et al., 2021; Gifford et al., 2013) and RNA-seq (Takashima et al., 2014).

f) Heatmap showing the frequency of SMAD2/3 peak centre locations with respect to ChromHMM states in naïve and primed hPSCs (Chovanec et al., 2021). SMAD2/3 peaks in naïve and primed hPSCs were annotated with their respective ChromHMM states. The annotations associated with the randomly-selected control regions reflect the overall genomic representation of chromatin states.

g-i) Density coloured scatter plots showing indicated ChIP-seq and ATAC-seq values (log_2 RPM) in naïve versus primed hPSCs. Each dot corresponds to one naïve-specific SMAD2/3 peak.

j) Differential motif enrichment reporting the top four motifs (ranked by p-value) at SMAD2/3 binding sites in naïve hPSCs that are enriched compared to motifs identified at SMAD2/3 binding sites in primed hPSCs.

Figure 3

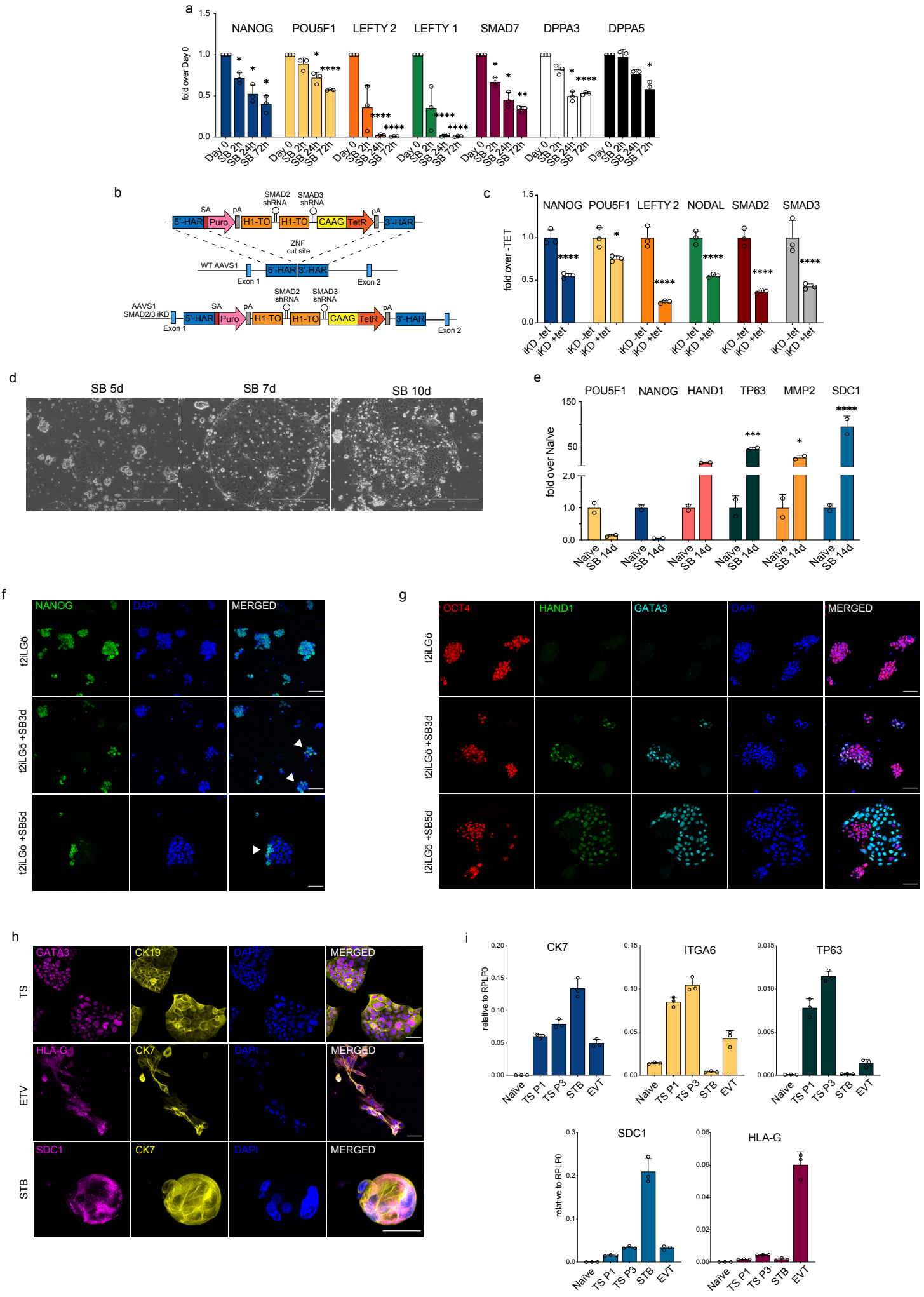


Figure 3. Inhibiting TGFβ signalling induces loss of pluripotency in naïve hPSCs

a) RT-qPCR expression analysis of pluripotency-associated genes and TGFβ-associated genes in naïve hPSCs (H9 NK2 line) following SB431542 treatment (t2iLGö + SB). Expression levels are shown as fold changes relative to day 0.

b) Schematic showing the integration of a single-step optimised inducible knock-down targeting construct into the *AAVS1* locus of H9 hPSCs, enabling the expression of SMAD2 and SMAD3 short hairpin RNAs (shRNAs) under the control of a tetracycline inducible promoter. ZFN: zinc-finger nucleases; 5'-HAR/3'-HAR: upstream/downstream homology arm; H1-TO: Tetracycline-inducible H1 Pol III promoter carrying one tet operon after the TATA box; CAAG: CMV early enhancer, chicken β-actin and rabbit β-globin hybrid promoter; TetR: Tetracycline-sensitive repressor protein; SA: splice acceptor; Puro, Puromycin resistance; pA, polyadenylation signal. Schematic adapted from (Bertero et al., 2016).

c) RT-qPCR analysis of gene expression levels in SMAD2/3 inducible knock-down (iKD) H9 naïve hPSCs following 5 days of tetracycline (tet) treatment. Expression levels are shown for each gene as fold change relative to iKD -tet. Cells were cultured in t2iLGö medium.

d) Phase contrast pictures of H9 NK2 naïve hPSCs after 5, 7, and 10 days of SB treatment in t2iLGö medium. Scale bars: 400 µm.

e) RT-qPCR analysis of trophoblast (*HAND1*, *TP63*, *MMP2* and *SDC1*) and pluripotency (*POU5F1*, *NANOG*) gene expression levels in naïve hPSCs following long term (14 days) SB treatment in t2iLGö medium. Expression levels are shown as fold changes relative to day 0 samples, n = 2 biological replicates.

f) Immunofluorescence microscopy showing the downregulation of NANOG (green) in naïve hPSCs following 3 and 5 days of SB treatment. DAPI signal in blue. White arrowheads indicate colonies displaying heterogeneous expression of NANOG. Scale bars: 50 µm.

g) Immunofluorescence microscopy for OCT4 (red), HAND1 (green), GATA3 (cyan) and DAPI (blue) in naïve hPSCs following 3 and 5 days of SB treatment in t2iLGö medium. Scale bars: 50 µm.

h) Immunofluorescence microscopy for GATA3, HLA-G, SDC1 (magenta), CK19 and CK7 (yellow), and DAPI (blue) in naïve-derived trophoblast stem cells (TS), extravillous trophoblast (ETV) and syncytiotrophoblast (STB). Scale bars: 50 µm.

i) RT-qPCR analysis of gene expression levels in naïve-derived trophoblast stem cells (TS), extravillous trophoblast (ETV) and syncytiotrophoblast (STB) compared to undifferentiated naïve hPSCs. Expression levels are shown for each gene relative to the housekeeping gene *RPLP0*.

RT-qPCR data show the mean ± SD of three biological replicates (unless specified otherwise) and were compared to their relative control using an ANOVA with Tukey's or Šídák's multiple comparisons test (*p ≤ 0.05, **p≤ 0.01, ***p≤ 0.001, ****p≤ 0.0001). See also Figure 3–Source Data 1.

Figure 4

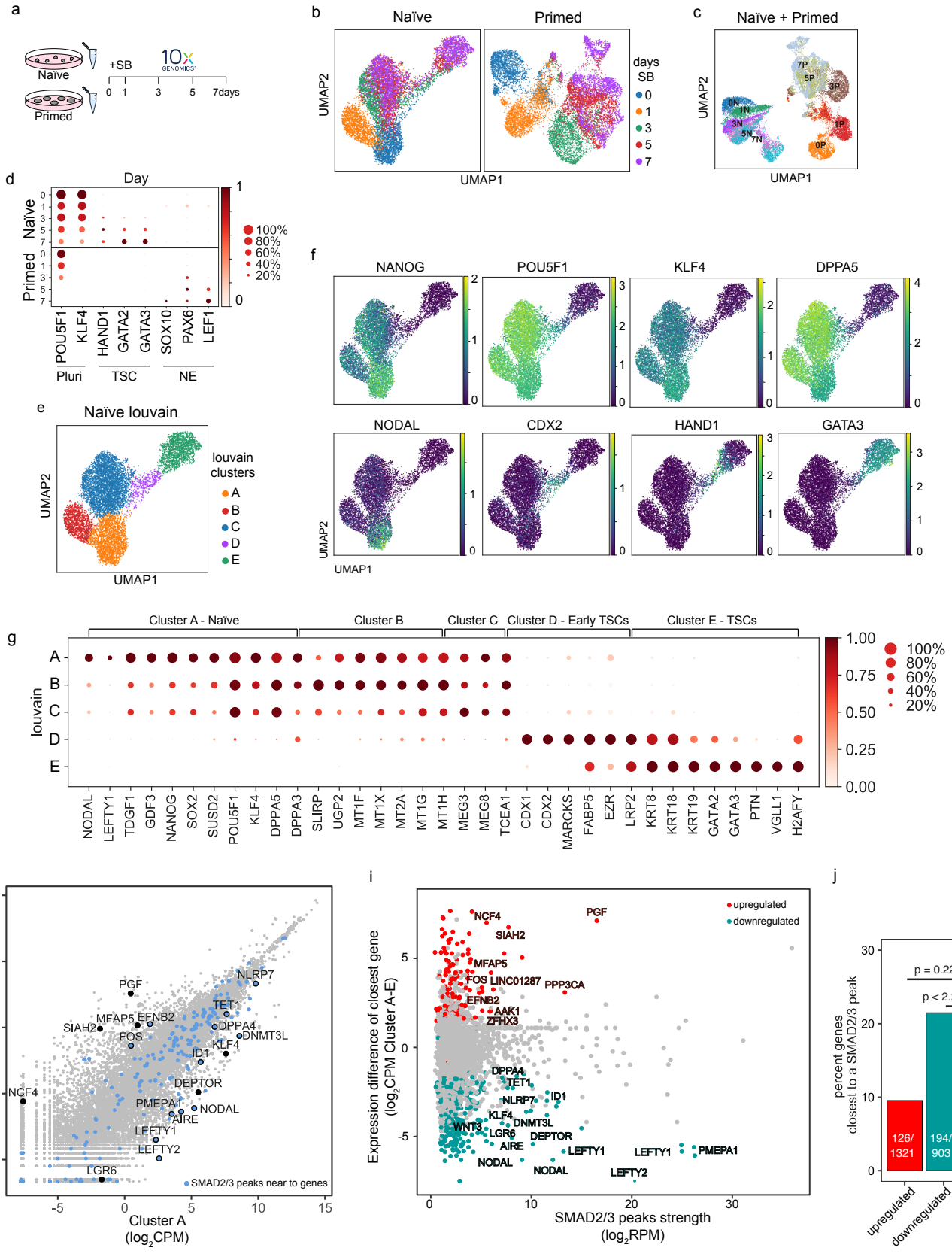


Figure 4. Single-cell transcriptional analysis reveals a trophoblast-like population arising in response to TGFβ inhibition in naïve hPSCs

a) Overview of the experimental procedure. Naïve and primed hPSCs were cultured in the presence of SB-431542 (SB), a potent TGFβ inhibitor, and samples were collected at days 0, 1, 3, 5 and 7. Single cell transcriptomes were obtained by 10X sequencing.

b) UMAP visualisation of naïve and primed cells during the SB time-course experiment, separated by days of treatment.

c) UMAP visualisation of the combined naïve and primed data set, separated by days of SB treatment (indicated by the number in the labels). N, naïve; P, primed.

d) Dot plot of selected gene expression values in naïve and primed cells during the SB time-course experiment, plotted by days of treatment (in rows). Each dot represents two values: mean expression within each category (visualised by colour) and fraction of cells expressing the gene (visualised by the size of the dot). Genes are indicative of pluripotent cells (Pluri), trophoblast stem cells (TSC) and neuroectoderm cells (NE).

e) UMAP visualisation of naïve hPSCs during the SB time-course experiment, separated by Louvain clustering (five clusters, A to E).

f) UMAP visualisation of naïve cells during the SB time-course experiment, showing the relative expression of pluripotency markers, *NANOG*, *POU5F1*, *KLF4*, and *DPPA5*; TGFβ effectors, *NODAL*; and trophoblast markers, *CDX2*, *HAND1*, *GATA3*.

g) Dot plot of expression values in naïve cells during the SB time-course experiment, separated by the five Louvain clusters. The genes shown represent a subset of the top 25 differentially expressed genes between the five clusters, as reported in Figure supplement 5e. Each dot represents two values: mean expression within each category (visualised by colour) and fraction of cells expressing the gene (visualised by the size of the dot).

h) Scatter plot reporting pseudobulk RNA-seq values (from 10X data) for cells in Louvain clusters A and E. Each dot represents one gene. Genes that have SMAD2/3 ChIP-seq peaks ($\log_2$ RPM > 5) within 12 kb of their transcription start site (TSS) are highlighted in blue and annotated. Several differentially expressed genes that are the closest gene to a SMAD2/3 peak (but are further away than 12 kb) are also named.

i) Scatter plot showing SMAD2/3 ChIP-seq peak strength ($\log_2$ RPM) versus the expression difference (cluster A – cluster E; $\log_2$ CPM) of the gene nearest to the SMAD2/3 peak. Upregulated genes, red; downregulated genes, green.

j) SMAD2/3 peaks were annotated with their nearest genes. Bar plot showing the percentage of genes that are the closest gene to a SMAD2/3 peak for genes that are upregulated (red) or

downregulated (green) between cells in clusters A and E. A randomly-selected set of control genes are shown in grey. The number of closest genes and the set size are reported within the bars. Statistical testing was performed using Chi-square test with Yates continuity and Bonferroni multiple testing correction.

Figure 5

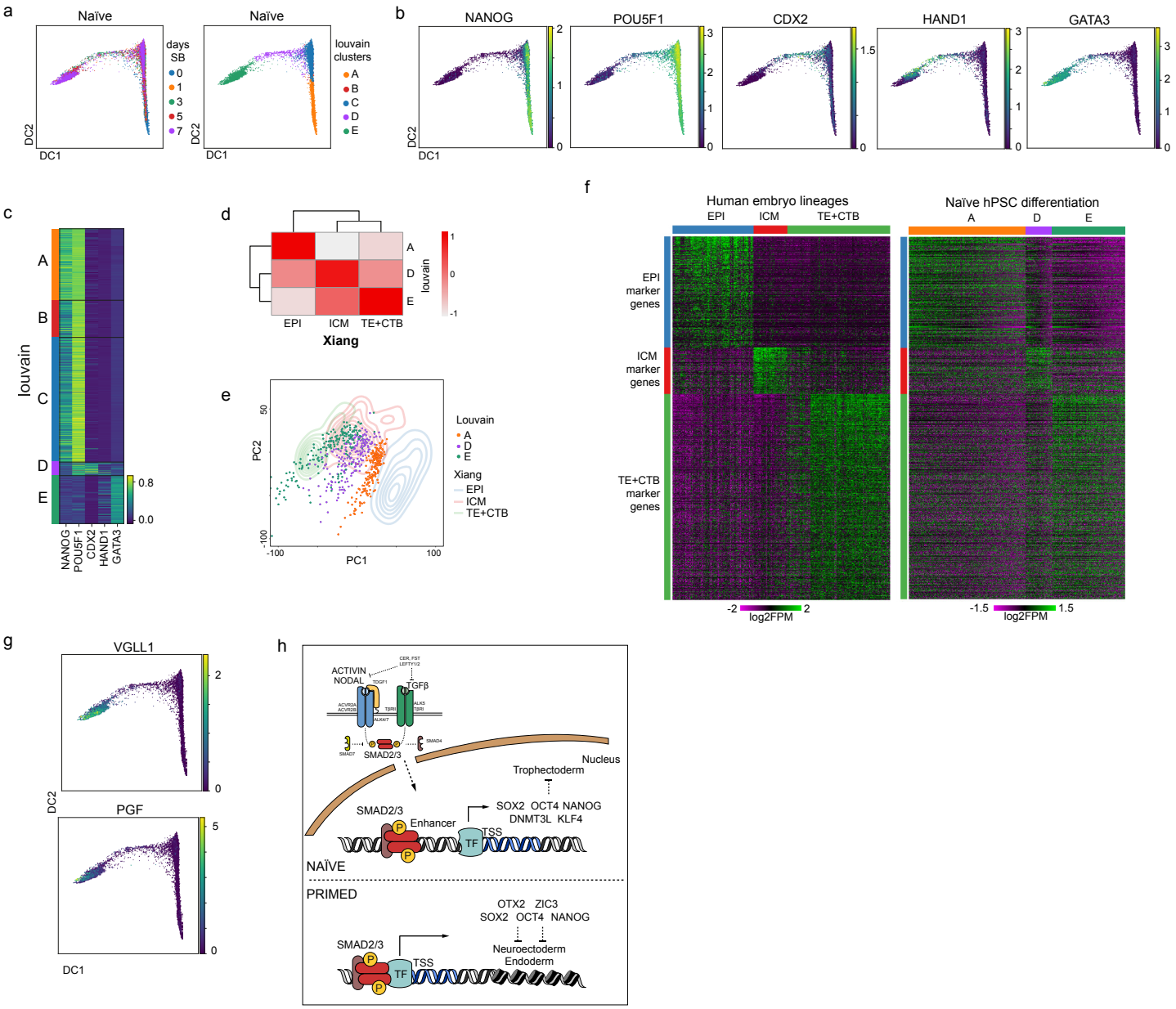


Figure 5. Differentiation of TGFβ-inhibited naïve hPSCs transcriptionally recapitulates early trophectoderm specification in human embryos

a) Diffusion maps of naïve cells during the SB time-course experiment, separated by days of treatment (left) and Louvain clustering (right).

b) Overlay of the diffusion maps with the relative expression of pluripotency markers *NANOG*, and *POU5F1*, and trophoblast markers *CDX2*, *HAND1*, *GATA3*.

c) Heatmap of the expression values of genes reported in Figure 5b separated by the Louvain clusters. Note the overlap in the expression of pluripotency and trophoblast markers in cells within cluster D.

d) Correlation plot between pseudobulk data from Louvain clusters A/D/E and EPI (Epiblast), ICM (Inner Cell Mass), and TE+CTB (Trophectoderm+Cytotrophoblast) from cultured human pre-gastrulation embryos (Xiang et al., 2020).

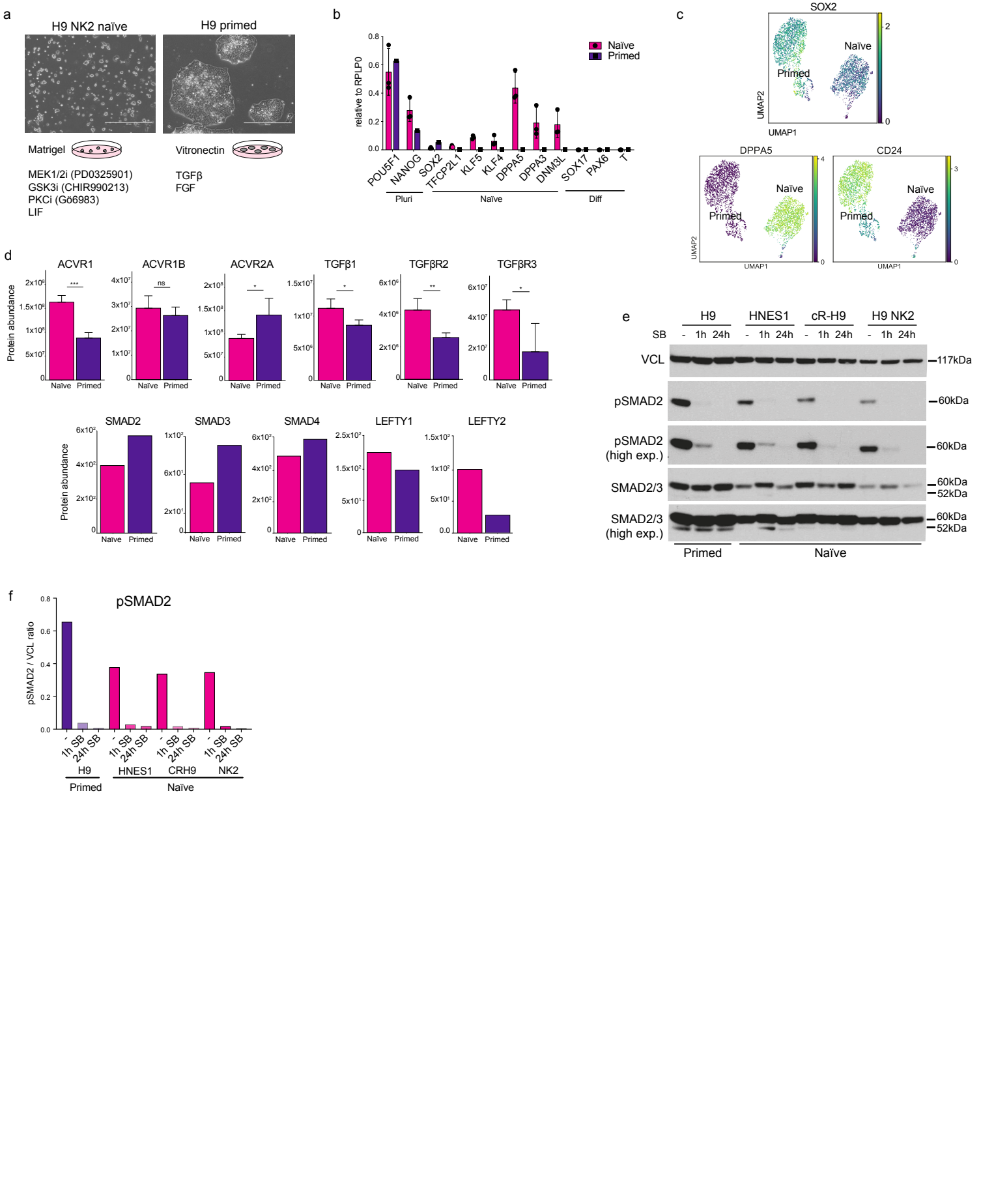
e) PCA plot overlapping 200 randomly-selected cells from each of the Louvain clusters A/D/E (individual dots) and data from 3D-cultured human pre-gastrulation embryos (Xiang et al., 2020), based on EPI, ICM and TE+CTB cells (contour lines). PC1 variance 2.15, PC2 variance 1.41.

f) Heatmaps visualising the expression of genes in EPI, ICM and TE+CTB (Xiang et al., 2020) and cells in Louvain clusters A/D/E. Note that the genes are in the same order for both plots.

g) Diffusion maps of naïve cells during the SB time-course experiment showing the relative expression of CTB markers – *VGLL1* and *PGF*.

h) We propose there is a continuum of TGFβ/Activin/Nodal signalling that spans a developmental window of human pluripotent states from naïve to primed. In both states, active TGFβ signalling promotes the expression of common pluripotency genes, such as *NANOG* and *POU5F1*, and contributes to the maintenance of pluripotency. SMAD2/3 are additionally required in naïve hPSCs to sustain the expression of naïve pluripotency factors, including *KLF4* and *DNMT3L*. Inactivating TGFβ signalling in naïve hPSCs leads to the downregulation of pluripotency genes, thereby enabling the induction of trophoblast differentiation.

Supplemental Figure 1



Supplemental Figure 1. Validation of naïve and primed hPSCs and TGFβ signalling pathway activation

a) Overview of the conditions used for culturing naïve and primed hPSCs and representative phase-contrast images.

b) RT-qPCR expression analysis of pan-pluripotency genes (Pluri), naïve markers (Naïve) and early-stage germ cell layer markers (Diff) in naïve and primed hPSCs. Data show the mean ± SD of three biological replicates, relative to the housekeeping gene *RPLP0*.

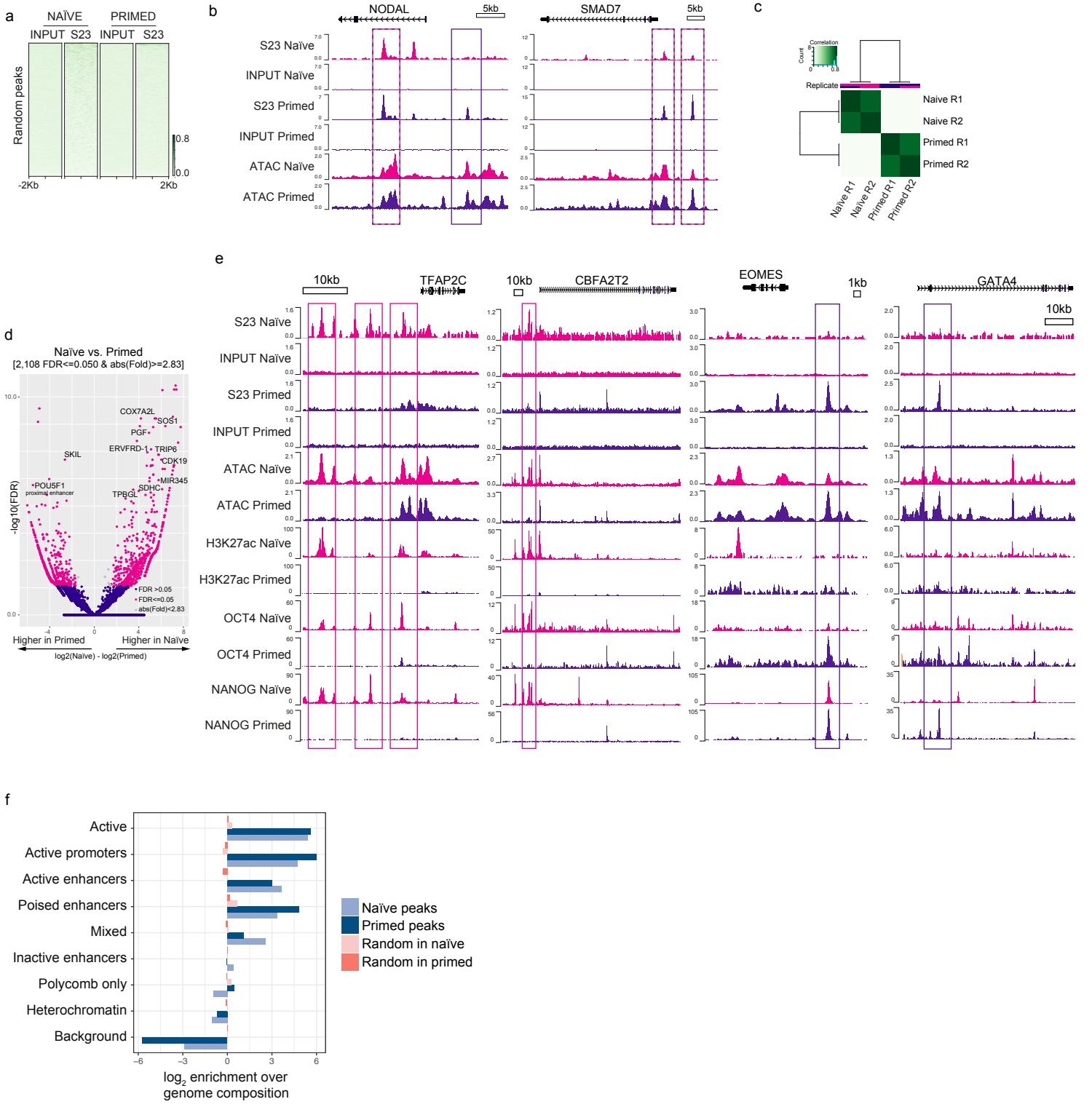
c) UMAP visualisation of naïve and primed cells reporting the relative expression of pluripotency marker *SOX2*, naïve marker *DPPA5*, and primed marker *CD24*.

d) Protein abundance levels for several TGFβ receptors (upper; (Wojdyla et al., 2020)) and transduction proteins (lower; (Di Stefano et al., 2018)) in naïve and primed hPSCs. The upper charts show the mean ±SD of three (naïve) or four (primed) biological replicates and were compared using a LIMMA-moderated t test with Benjamini-Hochberg correction (ns, $q > 0.05$, *$q < 0.05$, **$q < 0.01$, ***$q < 0.001$, ****$q < 0.0001$). The lower charts show protein abundance of one biological replicate per cell type.

e) Expanded western blot from Figure 1h comparing TGFβ pathway activation in primed H9 cells (cultured in E8 medium) and in three naïve hPSC lines cultured in t2iLGö medium: embryo-derived HNES1, chemically-reset cR-H9, and transgene-reset H9 NK2 naïve cells. Blots show SMAD2 phosphorylation signal and total SMAD2/3 in normal conditions (-), and following 1h and 24h of SB431542 supplementation to their culture media. Two separate exposures are shown. Vinculin (VCL) used as loading control.

f) Relative quantification of western blot from Figure 1h performed using the software Fiji (ImageJ) reporting the pSMAD2 / Vinculin (VCL) ratio.

Supplemental Figure 2



Supplemental Figure 2. SMAD2/3 binds to chromatin at common and state-specific sites

a) Heatmap displaying normalised SMAD2/3 ChIP-seq signal ±2kb from the centre of the peaks in a randomly-selected subset of regions, complementing Figure 2a.

b) Genome browser tracks overlapping SMAD2/3 (S23) binding sites in naïve and primed cells with chromatin accessibility (ATAC-seq) at the *NODAL* and *SMAD7* loci. Input is shown as control.

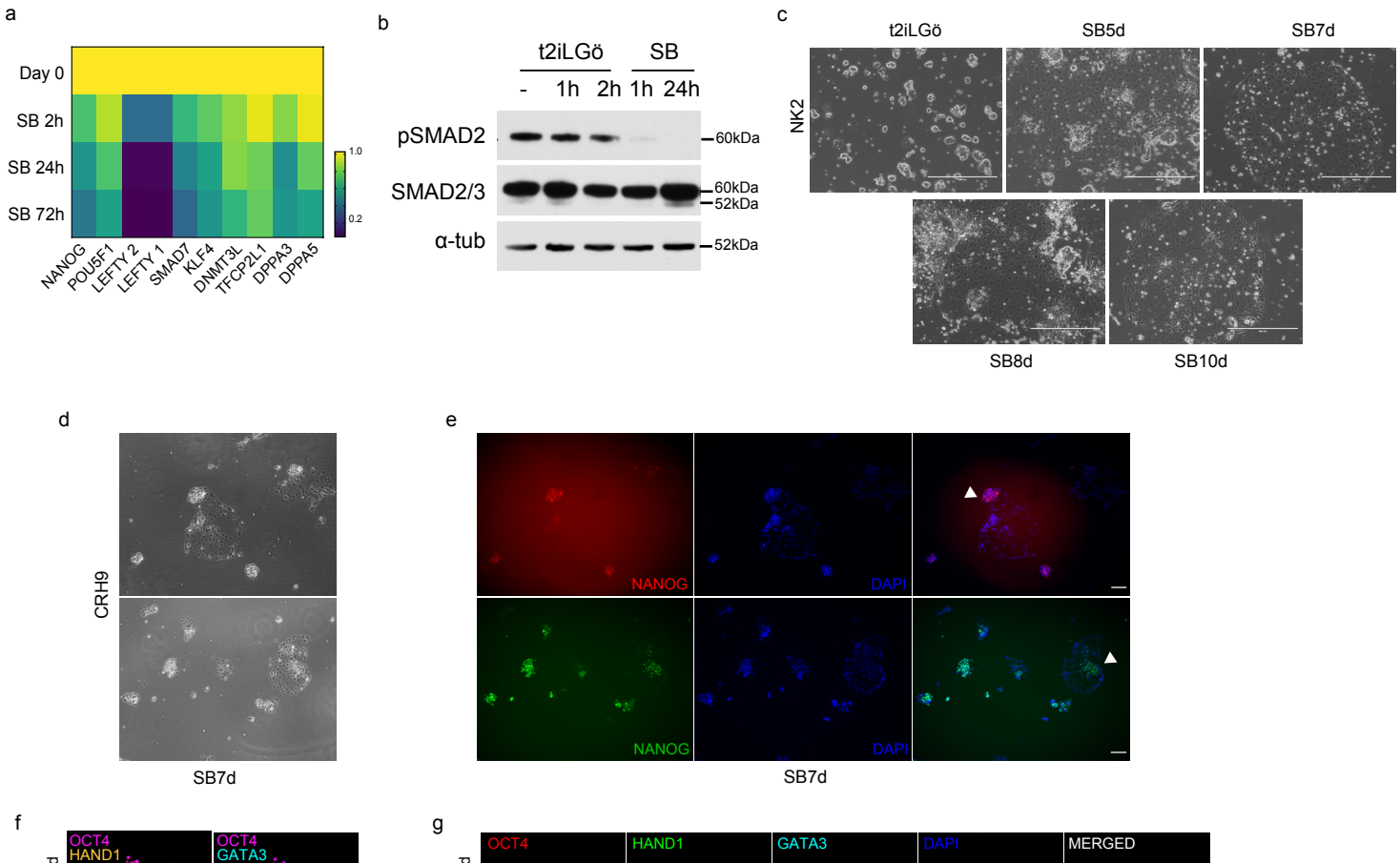
c) Correlation heatmap reporting the clustering of SMAD2/3 ChIP-seq replicates (R1 and R2) for each cell type, based on the count scores.

d) Volcano plot reporting the fold change in SMAD2/3 ChIP-seq signal at SMAD2/3 peaks between naïve and primed cells and the associated false discovery rate. Each dot represents a SMAD2/3 peak. Dots highlighted in pink are significantly differentially bound sites ($\log_{10}$ FDR ≤ 0.05; log2 fold changes ≥ 1.5).

e) Genome browser tracks exemplify the overlap between SMAD2/3 (S23) ChIP-seq binding sites in naïve and primed cells with chromatin accessibility (ATAC-seq), histone marks for active enhancers (H3K27ac), and OCT4 and NANOG ChIP-seq signal. Input is shown as control.

f) Bar plot reporting the log2 enrichment of SMAD2/3 peaks at ChromHMM-defined promoter and enhancer regions compared to other genomic regions. Related to Figure 2f.

Supplemental Figure 3



Supplemental Figure 3. TGFβ signalling inhibition induces loss of pluripotency in different naïve hPSCs

a) Heatmap reporting RT-qPCR expression analysis of pluripotency genes and TGFβ-associated genes in naïve cells following the addition of SB to the culture medium (t2iLGö), related to Figure 3a. Data show the mean of three biological replicates as fold changes compared to day 0.

b) Western blot showing TGFβ pathway activation in H9 NK2 naïve cells through the phosphorylation of SMAD2 (pSMAD2) and also total SMAD2/3 in normal conditions (-), after 1h and 2h of fresh media change (t2iLGö), and following 1h and 24h of SB treatment (t2iLGö+SB). Alpha-tubulin (α-tub) used as loading control.

c) Phase-contrast images of H9 NK2 naïve cells at day 0 (t2iLGö) and following 5, 7, 8 and 10 days of SB supplementation to the t2iLGö medium. Scale bars: 400µm.

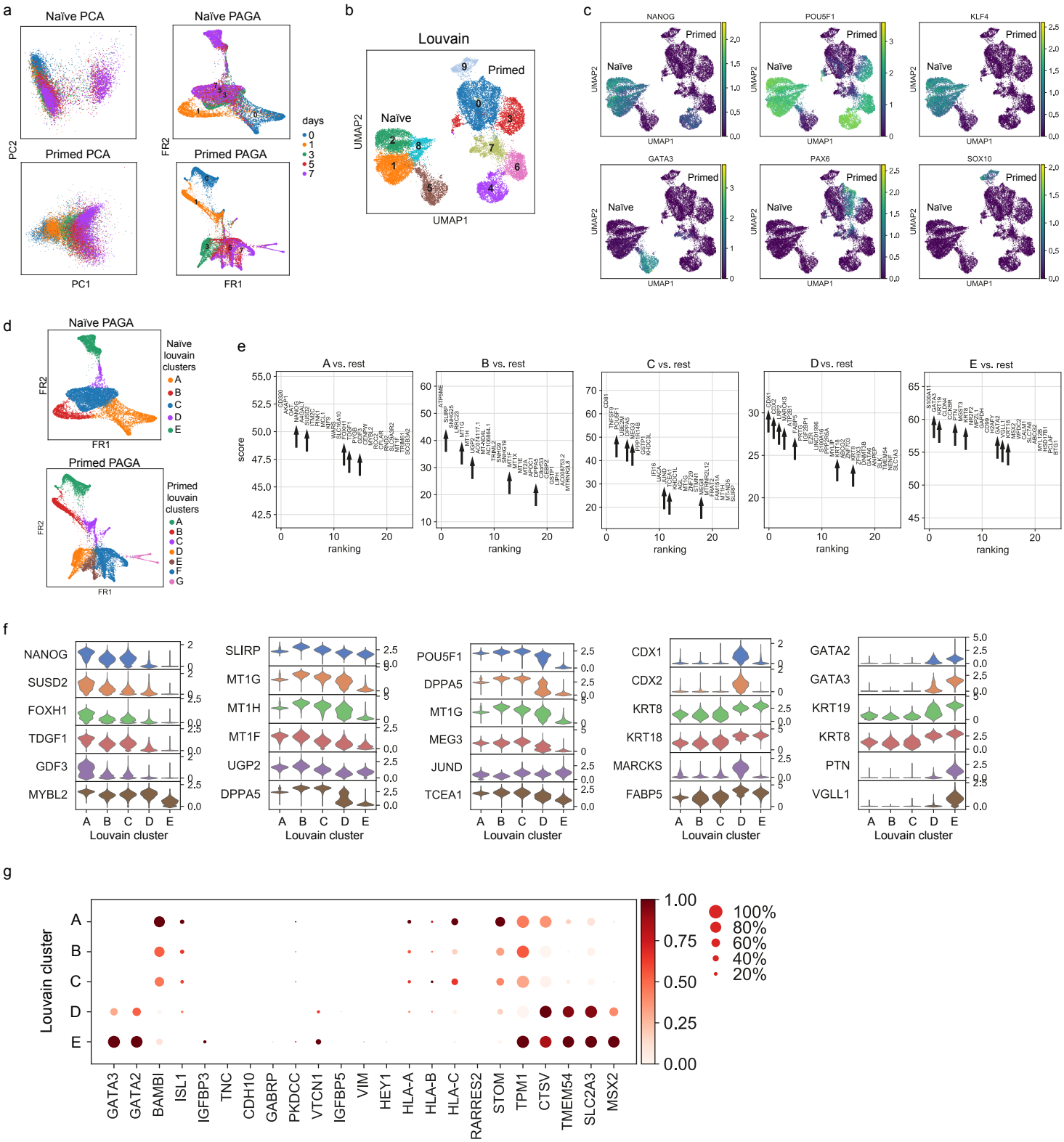
d) Phase-contrast images of chemically-reset cR-H9 naïve cells after 7 days of SB treatment.

e) Immunofluorescence microscopy of H9 NK2 naïve hPSCs for NANOG (red/green), and DAPI (blue), after 7 days of SB treatment in chemically-reset cR-H9 naïve cells, matching the phase-contrast images in Supplementary Figure 3d. White arrowheads indicate colonies displaying heterogeneous NANOG expression. Scale bars: 50 µm.

f) Immunofluorescence microscopy complementing Figure 3g showing (left) OCT4 (red) and HAND1 (green) overlap, or (right) OCT4 (red) and GATA3 (cyan) overlap after 3 and 5 days of SB treatment. Scale bars: 50 µm.

g) Immunofluorescence microscopy of H9 NK2 naïve hPSCs for OCT4 (red), HAND1 (green), GATA3 (cyan) and DAPI (blue) after 3 (upper) and 5 (lower) days of SB treatment. Scale bars: 50 µm.

Supplemental Figure 4



Supplemental Figure 4. Single-cell transcriptional analysis reveals different trajectories between naïve and primed hPSCs following TGFβ inhibition

a) PCA and partition-based graph abstraction (PAGA) connectivity visualisation of the 10X RNA-seq data in naïve and primed hPSCs during the SB time-course, separated by days of treatment.

b) UMAP visualisation of the combined naïve and primed cell 10X dataset during the SB time-course experiment, separated by Louvain clustering.

c) UMAP visualisation of the combined naïve and primed cell 10X dataset reporting the relative expression of pluripotency markers *NANOG*, *POU5F1*, *KLF4*, trophoblast marker *GATA3*, and neuroectoderm markers *PAX6*, *SOX10*.

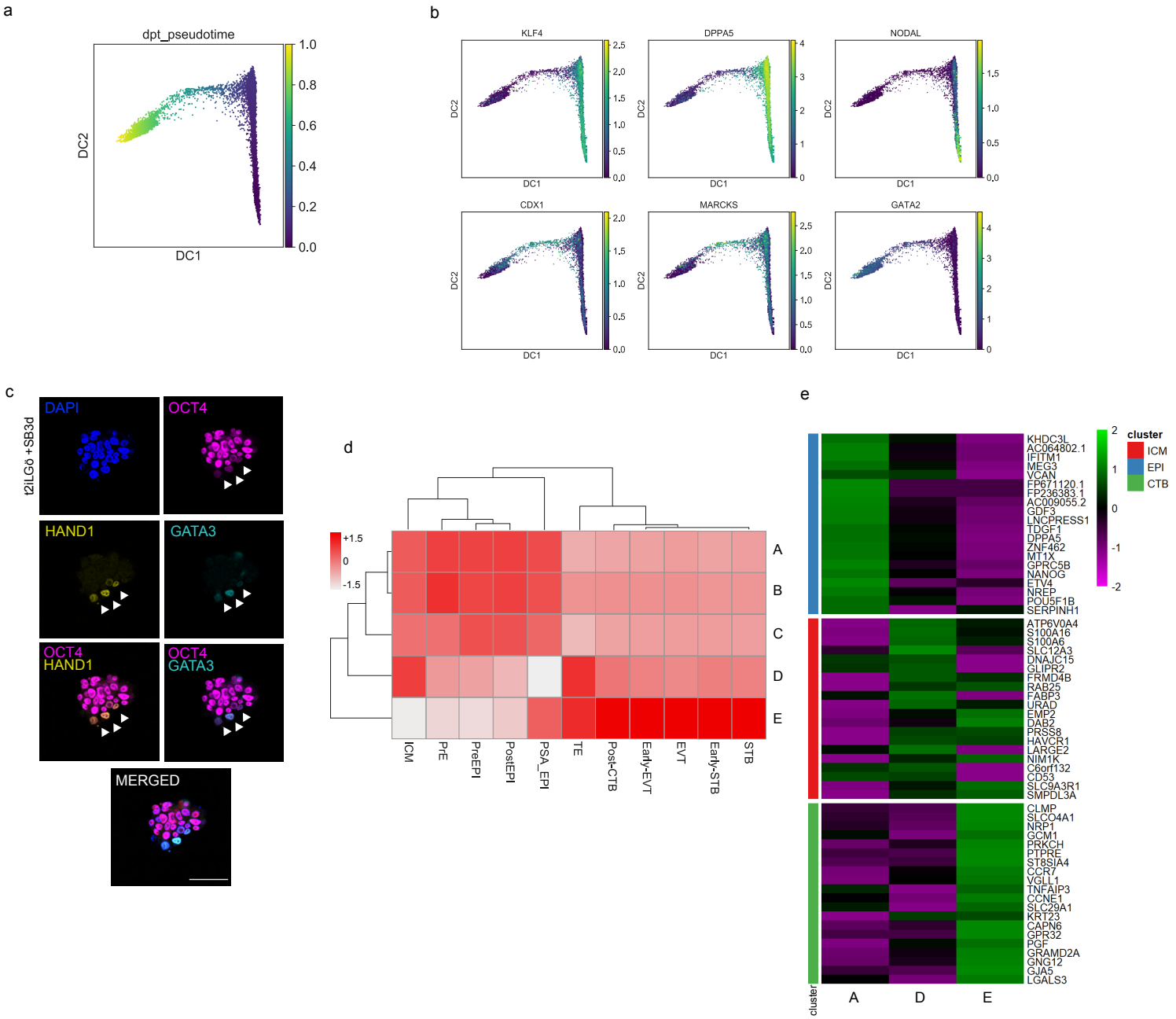
d) PAGA connectivity visualisation of naïve and primed cells during the SB time-course experiment, separated by Louvain clustering.

e) Summarised results showing the top 25 differentially expressed genes between Louvain clusters, identified by applying a Wilcoxon-Rank-Sum test. Black arrows highlight informative genes relative to each cluster.

f) Violin plots reporting the expression of a subset of genes identified in Figure 4g and Supplementary Figure 4e.

g) Dot plot of expression values in naïve cells during the SB time-course experiment, separated by the five Louvain clusters. The genes shown represent a subset of amnion marker genes reported in (Guo et al., 2021; Io et al., 2021; Zhao et al., 2021). Note that some genes are also expressed in trophoblast cells. Each dot represents two values: mean expression within each category (visualised by colour) and fraction of cells expressing the gene (visualised by the size of the dot).

Supplemental Figure 5



Supplemental Figure 5. Pseudotime trajectories of TGFβ-inhibited naïve hPSCs recapitulates early trophectoderm specification in human embryos

a) Diffusion maps of naïve hPSC 10X scRNA-seq data during the SB time-course, reporting diffusion pseudotime scores.

b) Diffusion map visualisation of naïve cells during the SB time-course reporting the relative expression of additional pluripotency markers *KLF4* and *DPPA5*, TGFβ effector *NODAL*, and trophoblast markers *CDX1*, *MARCKS* and *GATA2*.

c) Immunofluorescence microscopy for OCT4 (red), HAND1 (green), GATA3 (cyan) and DAPI (blue) in H9 NK2 naïve cells following 3 days of SB supplementation to t2iLGö medium. White arrowheads indicate cells that co-express low levels of OCT4 and HAND1 or GATA3. Scale bars: 50 µm.

d) Correlation heatmap between the pseudobulked datasets from all Louvain clusters (A/B/C/D/E) and the identified cell lineages from the cultured human pre-gastrulation embryos (Xiang et al., 2020). CTB: Cytotrophoblast; ETV: Extravillous trophoblast; STB: Syncytiotrophoblast; EPI: Epiblast; ICM: Inner Cell Mass; PrE: Primitive Endoderm, TE: Trophectoderm.

e) Heatmap reporting the expression in the Louvain clusters A, D, and E of the top 20 differentially expressed genes between EPI, ICM and CTB cell lineages (based on data published by (Xiang et al., 2020).

References

Alici-Garipcan A, Özçimen B, Süder I, Ülker V, Önder TT, Özören N. 2020. NLRP7 plays a functional role in regulating BMP4 signaling during differentiation of patient-derived trophoblasts. *Cell Death Dis* **11**:658.

Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**:532–538.

Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, Chambers I. 2018. Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell* **23**:276–288.e8.

Battle SL, Doni Jayavelu N, Azad RN, Hesson J, Ahmed FN, Overbey EG, Zoller JA, Mathieu J, Ruohola-Baker H, Ware CB, Hawkins RD. 2019. Enhancer Chromatin and 3D Genome Architecture Changes from Naive to Primed Human Embryonic Stem Cell States. *Stem Cell Reports* **12**:1129–1144.

Bayerl J, Ayyash M, Shani T, Manor YS, Gafni O, Massarwa R, Kalma Y, Aguilera-Castrejon A, Zerbib M, Amir H, Sheban D, Geula S, Mor N, Weinberger L, Naveh Tassa S, Krupalnik V, Oldak B, Livnat N, Tarazi S, Tawil S, Wildschutz E, Ashouokhi S, Lasman L, Rotter V, Hanna S, Ben-Yosef D, Novershtern N, Viukov S, Hanna JH. 2021. Principles of signaling pathway modulation for enhancing human naive pluripotency induction. *Cell Stem Cell*. doi:10.1016/j.stem.2021.04.001

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**:315–326.

Bertero A, Pawlowski M, Ortmann D, Snijders K, Yiangou L, de Brito MC, Brown S, Bernard WG, Cooper JD, Giacomelli E, Gambardella L, Hannan NRF, Iyer D, Sampaziotis F, Serrano F, Zonneveld MCF, Sinha S, Kotter M, Vallier L. 2016. Optimized inducible shRNA and CRISPR/Cas9 platforms forin vitrostudies of human development using hPSCs. *Development*. doi:10.1242/dev.138081

Bertero A, Yiangou L, Brown S, Ortmann D, Pawlowski M, Vallier L. 2018. Conditional Manipulation of Gene Function in Human Cells with Optimized Inducible shRNA. *Curr Protoc Stem Cell Biol* **44**:5C.4.1–5C.4.48.

Blakeley P, Fogarty N, Del Valle I, Wamaitha S, Hu TX, Elder K, Snell P, Christie L, Robson P, Niakan K. 2017. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Mechanisms of Development*. doi:10.1016/j.mod.2017.04.015

Bredenkamp N, Stirparo GG, Nichols J, Smith A, Guo G. 2019a. The Cell-Surface Marker Sushi Containing Domain 2 Facilitates Establishment of Human Naive Pluripotent Stem Cells. *Stem*

Cell Reports **12**:1212–1222.

Bredenkamp N, Yang J, Clarke J, Stirparo GG, von Meyenn F, Dietmann S, Baker D, Drummond R, Ren Y, Li D, Wu C, Rostovskaya M, Eminli-Meissner S, Smith A, Guo G. 2019b. Wnt Inhibition Facilitates RNA-Mediated Reprogramming of Human Somatic Cells to Naive Pluripotency. *Stem Cell Reports* **13**:1083–1098.

Brennan J, Lu CC, Norris DP, Rodriguez TA, Beddington RS, Robertson EJ. 2001. Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* **411**:965–969.

Brons IGM, Smithers LE, Trotter MWB, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**:191–195.

Brown S, Teo A, Pauklin S, Hannan N, Cho CH-H, Lim B, Vardy L, Dunn NR, Trotter M, Pedersen R, Vallier L. 2011. Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem Cells* **29**:1176–1185.

Cambuli F, Murray A, Dean W, Dudzinska D, Krueger F, Andrews S, Senner CE, Cook SJ, Hemberger M. 2014. Epigenetic memory of the first cell fate decision prevents complete ES cell reprogramming into trophoblast. *Nat Commun* **5**:5538.

Camus A, Perea-Gomez A, Moreau A, Collignon J. 2006. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev Biol* **295**:743–755.

Castel G, Meistermann D, Bretin B, Firmin J, Blin J, Loubersac S, Bruneau A, Chevolleau S, Kilens S, Chariau C, Gaignerie A, Francheteau Q, Kagawa H, Charpentier E, Flippe L, François-Campion V, Haider S, Dietrich B, Knöfler M, Arima T, Bourdon J, Rivron N, Masson D, Fournier T, Okae H, Fréour T, David L. 2020. Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells. *Cell Rep* **33**:108419.

Chan Y-S, Göke J, Ng J-H, Lu X, Gonzales KAU, Tan C-P, Tng W-Q, Hong Z-Z, Lim Y-S, Ng H-H. 2013. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell* **13**:663–675.

Chovanec P, Collier AJ, Krueger C, Várnai C, Semprich CI, Schoenfelder S, Corcoran AE, Rugg-Gunn PJ. 2021. Widespread reorganisation of pluripotent factor binding and gene regulatory interactions between human pluripotent states. *Nat Commun* **12**:2098.

Cinkornpumin JK, Kwon SY, Guo Y, Hossain I, Sirois J, Russett CS, Tseng H-W, Okae H, Arima T, Duchaine TF, Liu W, Pastor WA. 2020. Naive Human Embryonic Stem Cells Can Give Rise to Cells with a Trophoblast-like Transcriptome and Methylome. *Stem Cell Reports* **15**:198–213.

Collier AJ, Panula SP, Schell JP, Chovanec P, Plaza Reyes A, Petropoulos S, Corcoran AE, Walker R, Douagi I, Lanner F, Rugg-Gunn PJ. 2017. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. *Cell Stem Cell* **20**:874–890.

D'Amour KA, Agulnick AD, Eliazer S, Kelly OG, Kroon E, Baetge EE. 2005. Efficient

differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol* **23**:1534–1541.

Davidson KC, Mason EA, Pera MF. 2015. The pluripotent state in mouse and human. *Development* **142**:3090–3099.

Dawlaty MM, Ganz K, Powell BE, Hu Y-C, Markoulaki S, Cheng AW, Gao Q, Kim J, Choi S-W, Page DC, Jaenisch R. 2011. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**:166–175.

Di Stefano B, Ueda M, Sabri S, Brumbaugh J, Huebner AJ, Sahakyan A, Clement K, Clowers KJ, Erickson AR, Shioda K, Gygi SP, Gu H, Shioda T, Meissner A, Takashima Y, Plath K, Hochedlinger K. 2018. Reduced MEK inhibition preserves genomic stability in naive human embryonic stem cells. *Nat Methods* **15**:732–740.

Dong C, Beltcheva M, Gontarz P, Zhang B, Popli P, Fischer LA, Khan SA, Park K-M, Yoon E-J, Xing X, Kommagani R, Wang T, Solnica-Krezel L, Theunissen TW. 2020. Derivation of trophoblast stem cells from naïve human pluripotent stem cells. *Elife* **9**. doi:10.7554/eLife.52504

Dong C, Fischer LA, Theunissen TW. 2019. Recent insights into the naïve state of human pluripotency and its applications. *Exp Cell Res* **385**:111645.

Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, Kalma Y, Viukov S, Maza I, Zviran A, Rais Y, Shipony Z, Mukamel Z, Krupalnik V, Zerbib M, Geula S, Caspi I, Schneir D, Shwartz T, Gilad S, Amann-Zalcenstein D, Benjamin S, Amit I, Tanay A, Massarwa R, Novershtern N, Hanna JH. 2013. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**:282–286.

Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, Zhang X, Coyne M, Fostel JL, Holmes L, Meldrim J, Guttman M, Epstein C, Park H, Kohlbacher O, Rinn J, Gnirke A, Lander ES, Bernstein BE, Meissner A. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**:1149–1163.

Guo G, Stirparo GG, Strawbridge SE, Spindlow D, Yang J, Clarke J, Dattani A, Yanagida A, Li MA, Myers S, Özel BN, Nichols J, Smith A. 2021. Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell* **28**:1040–1056.

Guo G, von Meyenn F, Rostovskaya M, Clarke J, Dietmann S, Baker D, Sahakyan A, Myers S, Bertone P, Reik W, Plath K, Smith A. 2017. Epigenetic resetting of human pluripotency. *Development* **144**:2748–2763.

Guo G, von Meyenn F, Santos F, Chen Y, Reik W, Bertone P, Smith A, Nichols J. 2016. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports* **6**:437–446.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV,

Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**:108–112.

Inman GJ, Nicolás FJ, Callahan JF, Harling JD. 2002. SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Molecular Pharmacology* **62**:65–74.

Io S, Kabata M, Iemura Y, Semi K, Morone N, Minagawa A, Wang B, Okamoto I, Nakamura T, Kojima Y, Iwatani C, Tsuchiya H, Kaswandy B, Kondoh E, Kaneko S, Woltjen K, Saitou M, Yamamoto T, Mandai M, Takashima Y. 2021. Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell* **28**:1023–1039.

James D, Levine AJ, Besser D, Hemmati-Brivanlou A. 2005. TGFβ/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**:1273–1282.

Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, Weintraub AS, Hnisz D, Pegoraro G, Lee TI, Misteli T, Jaenisch R, Young RA. 2016. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18**:262–275.

Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, Rodig SJ, Daley GQ, Rao A. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**:200–213.

Linneberg-Agerholm M, Wong YF, Romero Herrera JA, Monteiro RS, Anderson KGV, Brickman JM. 2019. Naïve human pluripotent stem cells respond to Wnt, Nodal and LIF signalling to produce expandable naïve extra-embryonic endoderm. *Development* **146**. doi:10.1242/dev.180620

Liu X, Ouyang JF, Rossello FJ, Tan JP, Davidson KC, Valdes DS, Schröder J, Sun YBY, Chen J, Knaupp AS, Sun G, Chy HS, Huang Z, Pflueger J, Firas J, Tano V, Buckberry S, Paynter JM, Larcombe MR, Poppe D, Choo XY, O'Brien CM, Pastor WA, Chen D, Leichter AL, Naeem H, Tripathi P, Das PP, Grubman A, Powell DR, Laslett AL, David L, Nilsson SK, Clark AT, Lister R, Nefzger CM, Martelotto LG, Rackham OJL, Polo JM. 2020. Reprogramming roadmap reveals route to human induced trophoblast stem cells. *Nature* **586**:101–107.

Mahadevan S, Wen S, Wan Y-W, Peng H-H, Otta S, Liu Z, Iacovino M, Mahen EM, Kyba M, Sadikovic B, Van den Veyver IB. 2014. NLRP7 affects trophoblast lineage differentiation, binds to overexpressed YY1 and alters CpG methylation. *Hum Mol Genet* **23**:706–716.

Mathieu J, Ruohola-Baker H. 2017. Metabolic remodeling during the loss and acquisition of pluripotency. *Development* **144**:541–551.

Mesnard D, Guzman-Ayala M, Constam DB. 2006. Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development* **133**:2497–2505.

Niakan KK, Eggan K. 2013. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev Biol* **375**:54–64.

Okae H, Toh H, Sato T, Hiura H, Takahashi S, Shirane K, Kabayama Y, Suyama M, Sasaki H, Arima T. 2018. Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22**:50–63.

Pastor WA, Chen D, Liu W, Kim R, Sahakyan A, Lukianchikov A, Plath K, Jacobsen SE, Clark AT. 2016. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell* **18**:323–329.

Pastor WA, Liu W, Chen D, Ho J, Kim R, Hunt TJ, Lukianchikov A, Liu X, Polo JM, Jacobsen SE, Clark AT. 2018. TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat Cell Biol* **20**:553–564.

Petropoulos S, Edsgärd D, Reinius B, Deng Q, Panula SP, Codeluppi S, Reyes AP, Linnarsson S, Sandberg R, Lanner F. 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**:1012–1026.

Ralston A, Cox BJ, Nishioka N, Sasaki H, Chea E, Rugg-Gunn P, Guo G, Robson P, Draper JS, Rossant J. 2010. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* **137**:395–403.

Rossant J, Tam PPL. 2017. New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. *Cell Stem Cell* **20**:18–28.

Rostovskaya M, Stirparo GG, Smith A. 2019. Capacitation of human naïve pluripotent stem cells for multi-lineage differentiation. *Development* **146**. doi:10.1242/dev.172916

Sahakyan A, Kim R, Chronis C, Sabri S, Bonora G, Theunissen TW, Kuoy E, Langerman J, Clark AT, Jaenisch R, Plath K. 2017. Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* **20**:87–101.

Smith JR, Vallier L, Lupo G, Alexander M, Harris WA, Pedersen RA. 2008. Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev Biol* **313**:107–117.

Soncin F, Khater M, To C, Pizzo D, Farah O, Wakeland A, Arul Nambi Rajan K, Nelson KK, Chang C-W, Moretto-Zita M, Natale DR, Laurent LC, Parast MM. 2018. Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. *Development* **145**. doi:10.1242/dev.156273

Taei A, Rasooli P, Braun T, Hassani S-N, Baharvand H. 2020. Signal regulators of human naïve pluripotency. *Exp Cell Res* **389**:111924.

Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, Oxley D, Santos F, Clarke J, Mansfield W, Reik W, Bertone P, Smith A. 2014. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**:1254–1269.

Theunissen TW, Friedli M, He Y, Planet E, O'Neil RC, Markoulaki S, Pontis J, Wang H, Iouranova A, Imbeault M, Duc J, Cohen MA, Wert KJ, Castanon R, Zhang Z, Huang Y, Nery JR, Drotar J,

Lungjangwa T, Trono D, Ecker JR, Jaenisch R. 2016. Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19**:502–515.

Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, Fan ZP, Maetzel D, Ganz K, Shi L, Lungjangwa T, Imsoonthornruksa S, Stelzer Y, Rangarajan S, D'Alessio A, Zhang J, Gao Q, Dawlaty MM, Young RA, Gray NS, Jaenisch R. 2014. Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. *Cell Stem Cell* **15**:524–526.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**:1145–1147.

Touboul T, Hannan NRF, Corbineau S, Martinez A, Martinet C, Branchereau S, Mainot S, Strick-Marchand H, Pedersen R, Di Santo J, Weber A, Vallier L. 2010. Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology*. doi:10.1002/hep.23506

Vallier L, Alexander M, Pedersen RA. 2005. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118**:4495–4509.

Vallier L, Mendjan S, Brown S, Chng Z, Teo A, Smithers LE, Trotter MWB, Cho CH-H, Martinez A, Rugg-Gunn P, Brons G, Pedersen RA. 2009. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* **136**:1339–1349.

Vallot C, Patrat C, Collier AJ, Huret C, Casanova M, Liyakat Ali TM, Tosolini M, Frydman N, Heard E, Rugg-Gunn PJ, Rougeulle C. 2017. XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell* **20**:102–111.

Varlet I, Collignon J, Robertson EJ. 1997. nodal expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation. *Development* **124**:1033–1044.

Weinberger L, Ayyash M, Novershtern N, Hanna JH. 2016. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol* **17**:155–169.

Wojdyla K, Collier AJ, Fabian C, Nisi PS, Biggins L, Oxley D, Rugg-Gunn PJ. 2020. Cell-Surface Proteomics Identifies Differences in Signaling and Adhesion Protein Expression between Naive and Primed Human Pluripotent Stem Cells. *Stem Cell Reports* **14**:972–988.

Xiang L, Yin Y, Zheng Y, Ma Y, Li Y, Zhao Z, Guo J, Ai Z, Niu Y, Duan K, He J, Ren S, Wu D, Bai Y, Shang Z, Dai X, Ji W, Li T. 2020. A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**:537–542.

Xu R-H, Sampsell-Barron TL, Gu F, Root S, Peck RM, Pan G, Yu J, Antosiewicz-Bourget J, Tian S, Stewart R, Thomson JA. 2008. NANOG Is a Direct Target of TGFβ/Activin-Mediated SMAD Signaling in Human ESCs. *Cell Stem Cell* **3**:196–206.

Zhao C, Reyes AP, Schell JP, Weltner J, Ortega N. 2021. Reprogrammed iBlastoids contain amnion-like cells but not trophectoderm. *bioRxiv*. doi:10.1101/2021.05.07.442980