KLF17 promotes human naïve pluripotency but is not required for its establishment

Rebecca A. Lea¹, Afshan McCarthy¹, Stefan Boeing², Todd Fallesen³, Kay Elder⁴, Phil Snell⁴, Leila Christie⁴, Sarah Adkins⁵, Valerie Shaikly⁵, Mohamed Taranissi⁵, Kathy K. Niakan¹, ², ³

¹Human Embryo and Stem Cell Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
²Bioinformatics and Biostatistics Service, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
³Crick Advanced Light Microscopy, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
⁴Bourn Hall Clinic, Bourn, Cambridge, CB23 2TN, UK
⁵Assisted Reproduction and Gynaecology Centre, London, W1G 6LP, UK
⁶The Centre for Trophoblast Research, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK
*Correspondence: kkn21@cam.ac.uk

Summary statement
Investigating KLF17 in human pluripotency reveals that it is sufficient, but not necessary, to establish naïve hESCs. We posit that KLF17 is a peripheral regulator of human pluripotent stem cells.

Abstract
Current knowledge of the transcriptional regulation of human pluripotency is incomplete, with lack of inter-species conservation observed. Single-cell transcriptomics analysis of human embryos previously enabled us to identify transcription factors, including the zinc-finger protein KLF17, that are enriched in the human epiblast and naïve hESCs. Here we show that KLF17 is expressed coincident with the known pluripotency-associated factors NANOG and SOX2 across human blastocyst development. We investigate the function of KLF17 using primed and naïve hESCs for gain- and loss-of-function analyses. We find that ectopic expression of KLF17 in primed hESCs is sufficient to induce a naïve-like transcriptome and that KLF17 can drive transgene-mediated resetting to naïve pluripotency. This implies a role for KLF17 in establishing naïve pluripotency. However, CRISPR-Cas9-mediated knockout studies reveal that KLF17 is not required for naïve pluripotency acquisition in vitro. Transcriptome analysis of naïve hESCs identifies subtle effects on metabolism and signalling pathways following KLF17 loss-of-function, and possible redundancy with other KLF paralogues. Overall, we show that KLF17 is sufficient, but not necessary, for naïve pluripotency under the given in vitro conditions.

© 2021. Published by The Company of Biologists Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.
Key words:
Human embryonic stem cells, naïve pluripotency, epiblast, KLF17, KLF factors

Introduction
Model organisms such as the mouse have allowed molecular mechanisms that regulate early mammalian development to be identified (Rossant, 2016), some of which are conserved in humans (Gerri et al., 2020). Despite the continued importance of comparative studies in mouse and other organisms, some aspects of early development such as developmental timing, chromatin accessibility and transcription factor function are distinct compared to humans (Niakan and Eggan, 2013; Fogarty et al., 2017; Gao et al., 2018). In particular, the advent of single-cell sequencing technologies has allowed in-depth transcriptomic analysis of human embryos, revealing a number of molecular differences compared to the mouse (Yan et al., 2013; Blakeley et al., 2015; Petropoulos et al., 2016; Stirparo et al., 2018). Our previous analysis highlighted that several genes thought of as canonical pluripotency-associated factors in the mouse, including $KLF2$, $ESRRB$ and $BMP4$ (Blakeley et al., 2015), are not expressed in the pluripotent epiblast (EPI) of the human preimplantation embryo, which forms the embryo proper. Conversely, we also highlighted genes that are specifically enriched in the human EPI, but not expressed in the pluripotent cells of the mouse embryo, including transcriptional regulators and signalling components (Blakeley et al., 2015).

The zinc finger DNA-binding protein KLF17 is one of these human EPI-enriched genes. KLF17 belongs to the Krüppel-like transcription factor family involved in development, which includes KLF4, a commonly used reprogramming factor (Takahashi and Yamanaka, 2006) and KLF2, a known pluripotency regulator in the mouse (Hall et al., 2009). Given the lack of $KLF2$ expression in the human EPI, it is interesting to speculate that KLF17 might function in a similar way. Indeed, the expression patterns of $KLF2$ and $KLF17$ in the human embryo are diametrically opposite to those of $Klf2$ and $Klf17$ in the mouse embryo (Yan et al., 2013; Blakeley et al., 2015). While $Klf17$ appears to be maternally deposited in the mouse zygote and its expression is abolished around the 8-cell stage, $KLF17$ is dramatically upregulated in the 8-cell human embryo, following embryonic genome activation (EGA) (Yan et al., 2013; Deng et al., 2014; Blakeley et al., 2015). Conversely, $Klf2$ is expressed from the 2-cell stage, corresponding to mouse EGA, and continues through to the blastocyst stage, but human $KLF2$ is only expressed pre-EGA (Yan et al., 2013; Deng et al., 2014; Blakeley et al., 2015). The human $KLF17$ and $KLF2$ sequences share ~60% homology across the C-terminal region containing the functional $C_{2}H_{2}$-type zinc-finger domains. KLF17 and mouse KLF2 also have additional homologous regions (~50%) throughout the protein, including part of a region in mouse KLF2 annotated as a protein-protein interaction domain, which may contribute to regulation and/or functional specificity. Furthermore, in mouse embryonic stem cells (mESCs), the triple knockout of $Klf2$, $Klf4$ and $Klf5$ can be rescued by ectopic expression of human $KLF17$ or mouse $Klf17$ (Yamane et al., 2018). Finally, the human and mouse KLF17 protein sequences have less similarity overall than other pairs of KLF orthologues (van Vliet et al., 2006). This is all suggestive of rapid,
divergent evolution of the human and mouse KLF genes and a potential switching of their function between species.

To date, KLF17 has primarily been studied in the context of cancer, where it has been implicated as a tumour suppressor by its interaction with TGFβ/SMAD signalling (Ali et al., 2015b) and p53 (Ali et al., 2015a) and inhibition of epithelial-to-mesenchymal transition (Gumireddy et al., 2009; Zhou et al., 2016). Since the recognition of its human EPI-specific expression, KLF17 has been widely used as a marker of pluripotency in the human embryo (Blakeley et al., 2015; Guo et al., 2016; Shahbazi et al., 2017; Kilens et al., 2018). The expression of KLF17 throughout preimplantation development, and in particular in pluripotent cells, is also conserved in a number of other organisms, including non-human primates (rhesus monkey, Macaca mulatta (Wang et al., 2017); common marmoset, Callithrix jacchus (Boroviak et al., 2015); and cynomolgus monkey, Macaca fascicularis (Nakamura et al., 2016)), and pig (Sus scrofa (Bernardo et al., 2018; Ramos-Ibeas et al., 2019)). Intriguingly, KLF17 expression is not detectable in conventionally derived “primed” human embryonic stem cells (hESCs) (Blakeley et al., 2015; Stirparo et al., 2018), reflecting their post-implantation-like identity. However, newer methods for deriving and/or culturing hESCs and human induced pluripotent stem cells (hiPSCs) in a naïve pluripotent state result in the maintenance or reinstatement of KLF17 gene activity (Theunissen et al., 2014; Guo et al., 2017; Guo et al., 2016; Liu et al., 2017; Kilens et al., 2018). This pattern of expression suggests the intriguing possibility that KLF17 acts as a transcriptional regulator of human naïve pluripotency, as exhibited in the bona fide state of the preimplantation EPI and approximated in the in vitro naïve hESC models. This hypothesis has also been explored by independent transcriptome analysis (Stirparo et al., 2018).

Studies to date have conclusively shown only that KLF17 is a marker of human pluripotency. Here, we set out to determine the function of KLF17, finding that its induced expression in conventional hESCs is sufficient, alongside naïve-permissive pluripotency conditions, to induce a complete change in phenotype from primed to naïve pluripotency. However, we also find that the null mutation of KLF17 in conventional hESCs is not detrimental to naïve resetting or maintenance of the resulting naïve cells. Altogether, this suggests that KLF17 functions to regulate genes associated with human naïve pluripotency, but that there is a degree of redundancy in vitro, such that KLF17 itself is not strictly necessary for the acquisition and maintenance of naïve pluripotency.

Results

KLF17 expression in the human embryo is gradually restricted to the epiblast

Detailed single-cell RNA sequencing (scRNA-seq) studies highlight KLF17 as a molecular marker that is expressed in the human preimplantation EPI. First, we reassessed the protein expression dynamics of KLF17 in human embryos, to investigate its distribution across the developing blastocyst. We performed immunofluorescence (IF) analysis of KLF17 alongside the canonical pluripotency factors SOX2 and NANOG in human embryos from the early to late blastocyst stage (Figs. 1A, S1). NANOG
is the earliest known EPI-restricted factor in human embryos (Kimber et al., 2008; Niakan and Eggan, 2013), while the SOX2 expression dynamic closely resembles that of KLF17 (Blakeley et al., 2015). In keeping with previous data (Kilens et al., 2018), we found that in the earliest stage examined (early day 5 post-fertilisation (d.p.f)), KLF17 protein was detectable in the majority of cells of the embryo, with KLF17 expression in an average of 64% of all nuclei (Figs. 1, S1). Although the expression levels across all nuclei were heterogeneous, this widespread staining of KLF17 largely coincided with SOX2 in both the inner cell mass (ICM) and trophoectoderm (TE) populations at this stage (Figs. 1, S1). Indeed, an average of 70% of KLF17-positive cells at early day 5 d.p.f also expressed SOX2, compared to only 16.5% overlap with NANOG (Fig. 1B). As blastocyst development progressed, KLF17 expression was gradually restricted, with ICM enrichment by early day 6 d.p.f (Figs. 1, S1), as evidenced by the average overlap with SOX2 of 64% and NANOG of 54% (Fig. 1B). In early day 7 d.p.f blastocysts, KLF17 was restricted to the presumptive EPI cells, delineated by nearly exclusive co-staining with both SOX2 and NANOG (Figs. 1, S1). Interestingly, the restriction of KLF17 appeared to progress more slowly than that of SOX2. By late day 5 d.p.f, SOX2 was only appreciably expressed in the ICM and to a lesser extent in polar TE (mean 52% overlap with NANOG) and it was restricted to the NANOG-positive EPI in early day 6 d.p.f embryos (mean 78% overlap with NANOG) (Fig. 1C). In contrast, there remained appreciable KLF17 protein across cells of the TE in most of the late day 6 d.p.f blastocysts analysed (Figs. 1, S1). This suggests that the half-life of KLF17 protein may be longer than that of SOX2, given the absence of KLF17 transcripts in the extraembryonic lineages of human blastocysts by scRNA-seq analysis (Blakeley et al., 2015). As reported previously, NANOG was detected in relatively few cells at all stages of blastocyst development (Niakan and Eggan, 2013) (Fig. S1C). These NANOG-positive cells represent the preimplantation EPI. Despite the initial widespread expression pattern of KLF17, its gradual restriction to the NANOG/SOX2 dual-positive EPI suggests that it is specifically retained in the pluripotent compartment, perhaps to perform an unappreciated role in pluripotency regulation or EPI development.

**Induction of KLF17 promotes a naïve pluripotency-like phenotype in conventional hESCs**

Given that KLF17 is not expressed in conventional primed hESCs, we investigated the effect of ectopic overexpression of KLF17 in these conditions. We hypothesised that KLF17, as a transcriptional regulator that is enriched in the naïve state (Blakeley et al., 2015; Guo et al., 2016; Messmer et al., 2019), might be sufficient to regulate other naïve pluripotency-associated genes when ectopically expressed in primed pluripotent hESCs.

We generated hESCs with doxycycline (Dox)-inducible, 3’ HA-tagged KLF17 transgene expression (Fig. S2A) and found that 5 days treatment with 1 μg/ml Dox was sufficient for robust expression of KLF17 protein (Fig. S2B). We therefore examined the possibility of gene expression changes in response to ectopic KLF17 in primed culture conditions. Using quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR), we analysed the expression of a number of genes identified as either naïve- or primed-enriched through previous differential gene expression analyses (Stirparo et al., 2018; Messmer et al., 2019) after 5 days Dox induction (Fig. 2). We identified naïve-enriched
factors that were significantly upregulated in response to KLF17 induction: ARGFX (~65-fold; p=0.03), ZFP42 (~180-fold; p=0.02) DPPA5 (~3.9-fold; p=0.04), DNMT3L (~300-fold; p=0.003) and TFAP2C (~2-fold; p=0.03). Of these genes, our recent scRNA-seq analysis revealed that only ZFP42 is appreciably expressed in primed hESCs (Wamaitha et al., 2020). Therefore, expression of KLF17 alone is sufficient not only to upregulate a gene already active in conventional hESCs, but also to initiate expression of genes that are otherwise transcriptionally silent. On the other hand, expression of NANOG and endogenous KLF17 remained unchanged, revealing a lack of KLF17 auto-regulation (Fig. 2).

In order to understand the full extent of the gene expression changes following KLF17 overexpression, we performed mRNA-seq across a 5-day time course of induction. Dimensionality reduction by principal component analysis (PCA) separated the samples by treatment (uninduced (UI) or induced (+Dox)) and timepoint (Fig. 3A). The uninduced control cells progressed through PC2 with time, reflecting transcriptional changes that occur following passaging. However, even as early as day 2, the induced and uninduced hESCs were clearly separated by PC1 and PC2 (Fig. 3A). Thus, ectopic expression of KLF17 in primed hESCs was sufficient to rapidly bring about considerable transcriptome-wide changes.

To determine the nature of the genes impacted upon by KLF17, we performed differential gene expression analysis between uninduced and induced samples at each timepoint. At day 5, we uncovered 1760 and 1315 up- and downregulated genes, respectively (p_adj<0.05) (Fig. 3B; Table S1). Of the upregulated genes at day 5, 537 (31%) have been previously identified as enriched in naïve hESCs (Stirparo et al., 2018; Messmer et al., 2019) and/or the human EPI (Blakeley et al., 2015), including 46 genes that are EPI-enriched but not differentially expressed between naïve and primed hESCs (e.g., LEFTY1, CALB1, ETV5, ETV4 and PFKP) (Fig. 3C). In contrast, of the downregulated genes, 479 (36%) were previously identified as exclusively enriched in primed hESCs (e.g., ITGA2 and ITGA4, MAPK10, IGFBPL1 and IDO1) (Fig. 3D). This suggests that expression of KLF17 in hESCs cultured under primed culture conditions promotes a shift toward a more naïve pluripotent transcriptome. Indeed, this was supported by the observation that cells following one or two days Dox induction begin to cluster transcriptionally with bona fide resetting intermediates identified at day 10 during resetting driven by NANOG and KLF2 (NK2) expression (Collier et al., 2017) (Fig. 3E).

It thus appears that KLF17 alone is sufficient to induce significant transcriptional change in primed hESCs over 5 days. To identify those genes most likely to be directly regulated by KLF17, we performed a time course correlation analysis. Using a cut-off for the correlation coefficient of 0.85, we found 70 genes whose expression over time closely mimicked exogenous KLF17 (Fig. S3A; Table S2). Of these genes, two-thirds (47) were classified as significantly enriched from day 1 (p_adj<0.05) and almost all (69) were classified as significantly enriched from day 2 onwards (p_adj<0.05) (Fig. S3B-C), highlighting that these putative KLF17 targets were both rapidly and strongly upregulated following KLF17 induction. These genes included a number of components of the PI3K-AKT-mTOR signalling
pathway (*PIK3AP1*, *TSC2*, *NOS3*, *FGF18*, *FGFR3*, *ITGB7* and *LAMC2*; Fig. 3F), which is active in both primed and naïve hESCs and a driver of primed hESC and human EPI proliferation (Wamaitha et al., 2020). Following *KLF17* overexpression, hESCs significantly upregulated ligands, receptors and downstream components of the PI3K-AKT pathway (Fig. S4A-G), which are also enriched in the human EPI (Wamaitha et al., 2020). PI3K-AKT-mTOR signalling has also been implicated in an alternative state of naïve pluripotency (Duggal et al., 2015). This suggests that *KLF17* induction may modulate signalling through PI3K to a more naïve or EPI-like state.

To determine the activation state of the PI3K-AKT signalling pathway in uninduced and induced hESCs, we performed western blot analysis of the key players (Fig. S5). We found that phosphorylation of AKT was consistently decreased in response to *KLF17* expression at day 5 (Fig. S5C-D), as was phosphorylation of the more downstream effector, S6 (Fig. S5E). We also observed a decrease in phosphorylation of the upstream receptors, IGF1R and InsR (Fig. S5F), whereas phosphorylation of ERK1/2, which can crosstalk with AKT (Lamothe et al., 2004; Ornitz and Itoh, 2015), was consistently increased (Fig. S5G). The active, phosphorylated form of S6 is known to induce negative feedback at the level of upstream receptors (Harrington et al., 2004; Tremblay et al., 2007; Carracedo and Pandolfo, 2008). Meanwhile, pAKT(Serine473) is usually mediated by the activity of the downstream effector mTOR and stimulates full AKT activity (Alessi et al., 1996; Sarbassov et al., 2005), thereby regulating functions including metabolism, growth and proliferation. A reduction in the phosphorylation levels at various points in the pathway following upregulation of genes associated with PI3K-AKT-mTOR signalling may therefore indicate negative feedback, acting to keep the *KLF17*-induced hESCs in a steady state.

Other signalling factors were also highly correlated with *KLF17*, including *JAKMIP2*, *FGFR1* and *TNFRSF8* and the TGFβ signalling pathway components *LEFTY2* and *TGFB1I1*. Several cell adhesion-related and cytoskeletal proteins were also included in this list: *LAMC2*, *MUC4*, *COL5A1*, *ITGB7* and *MXRA5* (Fig. S4G-O). Given that changes in morphology and signalling are hallmarks of the conversion of primed to naïve pluripotency, it therefore appears that *KLF17* induces some of the same resetting-associated changes without any external signalling modulation.

Of note is the strong correlation of *KLF17* transgene expression and the long non-coding RNA *LINC-ROR* ($r = 0.921$), which is upregulated ~2.4-fold after 24hrs induction (Fig. S4P). *LINC-ROR* has been identified as a regulator of iPSC reprogramming (Loewer et al., 2010) and hESC self-renewal (Wang et al., 2013). Its expression is regulated by the core pluripotency transcription factors OCT4, SOX2 and NANOG (Loewer et al., 2010) and it in turn acts as a sink for pluripotency destabilising microRNAs that target the mRNA of these core factors for degradation (Wang et al., 2013). In this way, through increased *LINC-ROR* expression, ectopic *KLF17* may limit hESC differentiation.
Finally, we noted that terms related to WNT signalling were enriched amongst the 1711 genes downregulated after 24hrs of Dox induction (Table S3). Activity of the WNT pathway has been suggested to promote differentiation of hESCs, in both primed and naïve pluripotent states (Davidson et al., 2012; Singh et al., 2012; Bredenkamp et al., 2019b) and to be suppressed through crosstalk with the PI3K-AKT signalling pathway (Singh et al., 2012). Therefore, downregulation of genes associated with WNT signalling may suggest a mechanism through which KLF17-overexpressing hESCs would be refractory to differentiation cues.

Of the 50 WNT signalling-associated genes identified as significantly downregulated following 24hrs of \textit{KLF17} expression, 74% (37) have been suggested as sites of KLF17 binding under alternative naïve hESC culture conditions (Bayerl et al., 2021), including \textit{GSK3β}, \textit{TNKS2} and \textit{CTNNB1} (Fig. S6A-C). Similarly, of the 70 genes whose expression dynamics closely mimicked the \textit{KLF17} transgene, 63% (44) were identified as putative sites of KLF17 binding (Bayerl et al., 2021), including \textit{LAMC2} and \textit{FGF18} (Fig. S6D-E), suggesting that KLF17 may directly regulate the expression of these genes.

We confirmed the expression patterns of a number of differentially expressed genes (DEGs) by qRT-PCR (Fig. S6F-K) and/or immunofluorescence analysis (Fig. S6L), including \textit{DNMT3L}, \textit{VENTX}, \textit{GP130} and \textit{TFAP2C}, the latter of which is an essential regulator of naïve hESCs (Pastor et al., 2018). Altogether this supports our hypothesis that KLF17 acts to transcriptionally regulate genes associated with naïve human pluripotency.

\textbf{\textit{KLF17} expression drives hESCs to naïve pluripotency alongside signalling modulation}

Given that KLF17 is sufficient to upregulate naïve pluripotency-associated factors under conventional primed hESC conditions, we hypothesised that \textit{KLF17} induction may be sufficient to reset primed hESCs to a naïve pluripotent state under the appropriate culture regime. The use of ectopic gene expression to drive resetting is common, with deployment of transgenes including \textit{OCT4}, \textit{KLF4}, \textit{SOX2}, \textit{YAP}, \textit{NANOG} and/or \textit{KLF2} (Hanna et al., 2010; Theunissen et al., 2014; Takashima et al., 2014; Qin et al., 2016; Liu et al., 2017), with various media compositions.

Initial testing of \textit{KLF17} induction for 5 days under two naïve hESC culture conditions, tt2L+Gö (Guo et al., 2017) and PXGL (Bredenkamp et al., 2019a; Bredenkamp et al., 2019b), revealed considerably stronger expression of the naïve markers \textit{DNMT3L} and \textit{SUSD2} compared to cells treated equivalently in conventional (mTeSR1) medium or untreated controls (Figs. 4A, S7A). The upregulation of \textit{SUSD2} expression is particularly noteworthy, as it has been recently identified as a highly specific cell surface marker of naïve hESCs (Bredenkamp et al., 2019a). We attempted to propagate the cells, by single cell passaging, in both naïve and primed conditions after 5 days of \textit{KLF17} induction. Rounded and highly refractile colonies showing typical naïve hESC morphology began to appear only in cells treated with Dox in PXGL medium (Fig. 4B). Conversely, the uninduced cells in PXGL had largely died (Fig. S7B). While survival of uninduced cells was equally compromised following passaging in
tt2iL+Gö, the induced cells showed only evidence of differentiation, while all cells grown in mTeSR1 survived with typical primed hESC morphology (Fig. S7B). This suggests that the ectopic expression of KLF17 is sufficient to reset conventional hESCs to a naïve-like pluripotent phenotype when supported by PXGL medium (Guo et al., 2017). During chemical epigenetic resetting, PXGL medium supports the initial primed-to-naïve transition via WNT signalling modulation through XAV939 (Guo et al., 2017). This suggests that WNT inhibition is important alongside KLF17 overexpression for primed-to-naïve resetting.

Bulk, single-cell passaging of these naïve-like colonies allowed for stable propagation of KLF17-inducible naïve hESCs for a minimum of 5 passages in PXGL, without requiring additional transgene activation beyond the initial 5-day period of Dox treatment (Fig. 4B). We were able to confirm protein expression of naïve hESC markers, and of factors identified above as upregulated following KLF17 induction in primed culture conditions, although we also found that DNMT3L is no longer entirely nuclear in PXGL, perhaps due to the great increase in its level of expression (Fig. 4C).

Furthermore, we performed bulk mRNA-sequencing of cells undergoing KLF17-driven resetting (+Dox days 1, 2 and 5), alongside day 0 and uninduced controls and the KLF17-inducible naïve cells at passage 5 (p5) in PXGL. As a comparison we again incorporated the data representing the progression of primed to naïve hESCs driven by NK2 overexpression in t2iL+Gö conditions (Collier et al., 2017). PCA analysis revealed that while culture in PXGL alone for 48hrs causes considerable transcriptome-wide differences (Fig. 4D), uninduced cells regress to a more typical hESC transcriptome, as evidenced by clustering of these cells with the day 0 samples from this study and published primed hESCs (Collier et al., 2017) (Fig. 4D). In contrast, hESCs overexpressing exogenous KLF17 for 5 days in PXGL represent bona fide intermediates between the primed and naïve pluripotent states, given their clustering close to the H9 Nmin and N4+ intermediates of NK2-driven resetting (Collier et al., 2017) (Fig. 4D). Continued propagation of these cells to p5 in PXGL induced a somewhat similar naïve state to that previously reported (Collier et al., 2017). KLF17-inducible naïve hESCs at p5 are clearly well-separated in transcriptional space from conventional primed cells, but do not fall within the cluster of p5, p10 and established naïve hESCs in t2iL+Gö (Collier et al., 2017) (Fig. 4D).

To investigate the transcriptional changes that occur during KLF17-driven resetting in more detail, we performed DEG analysis between the induced and uninduced cells over time. In keeping with their proximity in the PCA, there were relatively few significant (p<0.05) DEGs at days 1 and 2. However, examination of the transcriptomes of KLF17-inducible hESCs following 5 days culture in PXGL revealed 5057 genes and 4405 genes significantly up- and downregulated in induced versus uninduced cells, respectively (Fig. 4E, Table S4). Of the most strongly and significantly enriched genes, we identified a number of factors previously highlighted as naïve-enriched genes (Stirparo et al., 2018; Messmer et al., 2019), including SUSD2, DNMT3L, ZFP42 and DPPA3, as well as genes previously associated with KLF17 overexpression in primed conditions, e.g., PIK3AP1 (Fig. 4E).
Conversely, some of the most significantly downregulated genes included a number of primed hESC markers (Stirparo et al., 2018; Messmer et al., 2019), e.g., PAX6, ZIC2, ZIC5, LGI1 and MAPK10 (Fig. 4E).

Examining the dynamics of gene expression in more detail, we manually defined three broad groups of genes—those that are specifically upregulated in response to KLF17 induction (Fig. S8A); those that are restrained or repressed in response to KLF17 induction (Fig. S8B); and those that change expression in response to the PXGL culture condition alone (Fig. S8C). While culture in PXGL initially leads uninduced hESCs towards a more naïve intermediate-like state (Figs. 4D, S8C), the specific transcriptional modulation of a number of factors either directly or indirectly by KLF17 is clearly required to enable primed-to-naïve-like transcriptional conversion (Fig. S8A-B). Interestingly, WNT pathway components are among those genes specifically restrained by KLF17 overexpression (Fig. S8B), reinforcing the notion of WNT signalling regulation via KLF17. We therefore demonstrated that KLF17 is a potent inducer of the naïve pluripotent state in hESCs, capable of synergising with the appropriate culture environment to bring about a switch from primed to naïve pluripotency.

**Designing a strategy for KLF17 mutation in hESCs**

Next, we sought to determine whether KLF17 expression is required for resetting of primed hESCs to naïve pluripotency. For this, we optimised a protocol for CRISPR-Cas9-mediated mutation of KLF17.

Using *in silico* tools, we designed five guide RNAs (gRNAs) against KLF17 (Figs. 5A, S9A). We introduced Cas9 and each gRNA in turn into primed hESCs and performed deep sequencing of the KLF17 on-target locus by MiSeq analysis. This revealed that the introduction of Cas9 and each of the gRNAs led to insertion and deletion (indel) mutations, with an average mutation efficiency of ~60% (Fig. 5B). However, gRNA KLF17(3_3) was clearly inferior and therefore we did not consider it any further.

In order to decide upon the optimal gRNA for generating KLF17 null mutant (KLF17−/−) hESCs, we investigated the nature of the indels resulting from CRISPR-Cas9 targeting in each case. Firstly, it was clear that both KLF17(1_1) and KLF17(1_2) were biased towards the introduction of very small indel mutations, with the vast majority of indels less than 10 bp in size (Fig. S9B). Targeting with either of the exon 1-targeted gRNAs would thus leave the possibility of KLF17 expression from an identified alternative initiating methionine, with the possibility of generating a hypomorph with unexpected consequences. We therefore focused on the exon 3-targeting gRNAs, KLF17(3_1) and KLF17(3_2). The overall efficiency of these two gRNAs was very similar, but sequence analysis revealed a stronger propensity for the introduction of larger frameshift alleles by KLF17(3_1) (Fig. 5C). By disrupting a larger stretch of sequence within the region encoding the KLF17 DNA-binding domain, longer frameshift indels would be expected to lead to null mutations. We therefore determined to generate KLF17−/− hESC lines using gRNA KLF17(3_1).
**KLF17<sup>-/-</sup> hESCs are not impaired in their ability to adopt naïve pluripotency**

Following nucleofection and single-cell amplification of wild-type, primed hESCs, we generated 8 KLF17-targeted clones (Fig. S10A). Initial genotyping by short-range PCR and next-generation MiSeq suggested a high proportion of homozygous editing (5 of 8 edited clones; Fig. S10A-B). However, analysis of a ~950 bp region surrounding the on-target site revealed that these apparent homozygous clones had actually undergone an unexpected, long-range editing event on one allele (Fig. S10A-B). This was apparent from the lack of amplification of both alleles, as determined by the presence of only one variant-type at a highly polymorphic site in the human genome, while the remaining wild-type and heterozygous clones confirmed that this variant is heterozygous in the parental cells (Fig. S10A-C).

The extent of the damage was only determined in one clone, #9, where a 163 bp deletion could be detected in the sequence. For the remaining 4 clones, the damage apparently completely prevented amplification of the second allele. This highlights the importance of in-depth genotyping following CRISPR-Cas9-mediated mutagenesis, as previously noted (Kosicki et al., 2018; Cullot et al., 2019; Rayner et al., 2019; Przewrocka et al., 2020; Alanis-Lobato et al., 2021). We therefore sought to test whether clones #18 and #19, compound mutants with two frameshifted alleles predicted to introduce premature stop codons in the sequence encoding the third zinc fingers (Fig. S10D-F), were null for KLF17 expression.

We subjected three wild-type control clones and clones #18 and #19 to chemical resetting (Guo et al., 2017) for eight days. In control cells, we observed robust coexpression of KLF17 and OCT4, while the compound mutants lacked detectable KLF17 protein (Fig. 5D). To determine whether KLF17<sup>-/-</sup> hESCs were able to adopt a naïve pluripotent state, we repeated the chemical resetting and found that both the wild-type controls and KLF17<sup>-/-</sup> hESCs could be propagated in tt2iL+Gö conditions for at least 10 passages, maintaining typical naïve morphology (Fig. 6A). To identify molecular differences arising in KLF17<sup>-/-</sup> hESCs, we performed mRNA-seq at various timepoints throughout the chemical resetting process (Fig. S11A). The lack of appreciable KLF17 RNA expression (TPM<5) in the compound mutant clones (Fig. 6C) suggests that the presence of premature termination codons following the CRISPR-Cas9 target site induced nonsense-mediated decay of the mRNA during translation (Nickless et al., 2017). Clones #18 and #19 are therefore bona fide KLF17-null mutant hESCs. Despite this, PCA analysis of all samples revealed tight clustering of wild-type and KLF17<sup>-/-</sup> hESCs at all timepoints (Fig. 6B), and with previously published data (Collier et al., 2017) (Fig. S11B). This is consistent with the fact that the KLF17<sup>-/-</sup> cells were able to reset and survive long term under naïve culture conditions. This may indicate that KLF17 expression is not required for resetting under the given conditions, or there may be redundancy with other genes that compensate for null mutations in KLF17.

Interestingly, DESeq2 analysis identified the KLF17 parologue KLF5 as an early DEG, being significantly upregulated in KLF17<sup>-/-</sup> versus wild-type naïve hESCs at day 2 of the resetting process (Fig. 6D), a timepoint when KLF2 and KLF4 are not differentially expressed (Fig. 6E-F). Furthermore, both KLF5 and KLF4 are significantly (p<0.05) upregulated following culture of KLF17<sup>-/-</sup> hESCs in...
naïve conditions up to p5 (Fig. 6D,F), suggesting possible compensation by one or more KLF paralogues.

Despite this possible redundancy and the clear lack of overt phenotype in $KLF17^{-/-}$ naïve hESCs, further analysis revealed an increase in the number of DEGs between wild-type and mutant hESCs at naïve p5, by which point the naïve pluripotent phenotype is suggested to become more stable (Guo et al., 2017). At p5, 316 genes were significantly upregulated and 311 genes were significantly downregulated ($p_{adj}<0.05$) (Fig. 6G; Table S5). Among the genes most significantly downregulated was the RNA-binding protein LIN28A (Peng et al., 2011), which was persistently downregulated from p5 onwards (Fig. 6H). LIN28A has been implicated in pluripotency regulation (Heo et al., 2008; Kim et al., 2014; Viswanathan et al., 2008; Yu et al., 2007), though a potential role specifically in naïve hESCs has not been explored. Nevertheless, its significant and maintained downregulation may suggest that its expression depends either directly or indirectly on KLF17 in tt2iL+Gö conditions.

A number of rate-limiting enzymes of glycolysis were also significantly downregulated at p5, including HK2, PFKL, ENO1, ENO2, PGK1 and PKM (Fig. S11C-H). Moreover, WNT ligands, receptors and scaffolding proteins were upregulated in the mutant cells at p5 (Figs. 6I, S11I-J; Table S6). Unlike LIN28A, however, expression of these transcripts had recovered at later timepoints, altogether supporting the conclusion that expression of KLF17 is not required for the conversion of primed to naïve hESCs.

Discussion

In this study, we investigate the human EPI-enriched transcription factor KLF17. By IF analysis of developing human blastocysts, we show that the protein dynamics of KLF17 are remarkably similar to those of the known pluripotency-associated factor SOX2. Both transcription factors display widespread expression in the early blastocyst, with gradual restriction to the pluripotent EPI, marked by NANOG expression. This protein expression dynamic is also shared by the core pluripotency regulator OCT4 (Niakan and Eggan, 2013). Therefore, the expression pattern of KLF17 during preimplantation human development is suggestive of a role in pluripotency regulation.

Indeed, we show that KLF17 can induce the expression of a naïve hESC-like transcriptome in primed hESCs and is sufficient for primed-to-naïve hESC conversion. This implies that KLF17 is a powerful inducer of the human naïve pluripotent state in vitro. A previous study highlighted that both KLF17 and KLF4 can regulate gene expression through enhancers located within transposable elements (Pontis et al., 2019), which are often situated near to genes involved in embryo genome activation. KLF17-overexpressing hESCs were found to upregulate expression of transposons associated with naïve hESCs, leading to induction of genes involved in mitochondrial function, WNT signalling, cell cycle, adhesion and polarity (Pontis et al., 2019). This is in keeping with our transcriptome analysis, which suggests that KLF17 may be involved in direct regulation of various signalling pathways,
primarily PI3K-AKT and WNT, with delayed and indirect regulation of naïve pluripotency-associated markers like DNMT3L and SUSD2. Given the roles of PI3K-AKT (Wamaitha et al., 2020) and WNT signalling in human pluripotency and differentiation (Singh et al., 2012; Bredenkamp et al., 2019b; Mathieu et al., 2019), and the importance of WNT inhibition for recent methods of naïve pluripotency establishment (Zimmerlin et al., 2016; Guo et al., 2017; Bredenkamp et al., 2019b), we hypothesise that KLF17 acts to endogenously dampen WNT signalling in order to promote naïve pluripotency and inhibit pro-differentiation cues.

Nonetheless, we found that KLF17-driven resetting was successful only in PXGL medium but not in tt2iL+Gö. This suggests a requirement for exogenous WNT inhibition, thus implying that the effect of KLF17 expression on WNT signalling activity may not be sufficient alone, or sufficiently rapid, to enable the primed-to-naïve transition. This contrasts with KLF4, which was found to be sufficient for such conversion in t2iLGöY medium (Liu et al., 2017). Differences in the timing and levels of exogenous expression between the Sendai and mRNA methods used for KLF4 (Liu et al., 2017) and the Dox-inducible overexpression system used for KLF17 (this work) might account for this discrepancy. Alternatively, it may suggest that KLF4 is a more potent inducer of the naïve state of human pluripotency. Specifically, it will be interesting in the future to determine if KLF4 also acts by dampening the expression of WNT signalling components and in this way, bypasses a requirement for exogenous WNT inhibition.

Nevertheless, we also find that loss of KLF17 function is not detrimental to hESC resetting. This is surprising, given the rapid upregulation of KLF17 expression that has been reported during chemical resetting (Guo et al., 2017) and raises the possibility of genetic compensation. Indeed, KLF2, KLF4 and KLF5, which have all been implicated in pluripotency regulation and to have redundant functions with human KLF17 in mESCs (Yamane et al., 2018), are also rapidly upregulated in the early stages of resetting in both wild-type and KLF17−/− hESCs. Furthermore, we observe upregulation of KLF5 during the early stage of resetting, where hESCs are undergoing global epigenetic “opening” in response to histone deacetylase inhibition (Guo et al., 2017), and again at p5. This suggests that expression of KLF5 may be sufficient to compensate for a function carried out by KLF17 in the wild-type state.

In human embryo development, the localisation (Fogarty et al., 2021) and expression dynamics of KLF5 and KLF17 are highly correlated (r = 0.76; (Yan et al., 2013)), whereas the correlation coefficient of KLF4 (r = 0.48) is lower and KLF2 is not expressed in the human pluripotent EPI (Blakeley et al., 2015). Additionally, in vitro evidence points to overlapping functions at the molecular level. For instance, overexpressing KLF5 in mESCs increases self-renewal through specific upregulation of the AKT coactivator Tcl1 (Ema et al., 2008). In the present study, we identified a positive correlation between the expression of exogenous KLF17 and several PI3K-AKT pathway components and, indeed, the paralogue TCL1B is significantly upregulated following 5 days induction of KLF17 expression (Table S1). Further work could address this question of compensation by
performing dual knockout of both KLF17 and KLF5 in hESCs and investigating the cells’ competency to undergo chemical resetting.

Alternatively, other KLF paralogues, such as KLF4, may functionally compensate (Yamane et al., 2018), or the combinatorial action of numerous naïve hESC-associated transcription factors with overlapping targets might maintain KLF17-null naïve hESCs. This would establish KLF17 as a “peripheral” regulator of human pluripotency and suggest that KLF17, while individually dispensable, is able to reinforce the stability of the pluripotent state mediated by the core factors, OCT4 and SOX2 (Nichols and Smith, 2012). For instance, knockdown of either Klf2, Klf4 or Klf5 in naïve mESCs does not appear detrimental (Jiang et al., 2008; Yamane et al., 2018) and Klf2- or Klf4-null mutant embryos are viable through preimplantation development (Wani et al., 1998; Ehlermann et al., 2003). Despite this, all three factors have validated roles in pluripotency (Parisi et al., 2008; Hall et al., 2009; Jiang et al., 2008). This may suggest that only combinatorial mutation of the KLF factors would be sufficient to induce a detrimental phenotype in naïve hESCs, which will be interesting to explore in the future.

While KLF17−/− naïve hESCs did not overtly differ from wild-type counterparts, we did find interesting trends by differential gene expression at p5 of naïve culture, when WNT inhibition by XAV939 is withdrawn. We observe significant downregulation of metabolism and translation, concomitant with upregulation of protein degradation, in KLF17−/− naïve hESCs. This could reflect cellular stress reminiscent of proteasomal inhibition of primed hESCs (Saez et al., 2018), or could result from misexpression of specific genes such as the RNA-binding protein LIN28A, which has been directly implicated in the growth and survival of hESCs (Peng et al., 2011). Interestingly, LIN28 has been recently identified as a naïve-specific marker in porcine ESCs (Chen et al., 2020), at both the RNA and protein level, suggesting that its downregulation may be somewhat detrimental also in naïve hESCs. In the future it will be interesting to understand if the transcriptional changes following loss of KLF17 indicate induction of a cellular stress response or if there may be further post-transcriptional or translational effects that are not reflected in current transcriptional analysis.

Overall, our overexpression studies show that KLF17 may typically have a role in regulating naïve pluripotency in vitro. Nonetheless, it is clear from our data that KLF17 expression is not necessary for establishing naïve hESCs via chemical resetting (Guo et al., 2017). However, the effect of KLF17 loss in established naïve hESCs, or the human preimplantation embryo, remains unexplored. Furthermore, a functional requirement for KLF17 expression has been demonstrated under alternative culture conditions for naïve-like hESCs (Bayerl et al., 2021), suggesting that its relative importance may be context-specific. We therefore theorise that in a wild-type situation, KLF17 may act as a peripheral pluripotency factor in human naïve pluripotency, acting alongside a core pluripotency network of OCT4 and SOX2 to maintain robustness of the pluripotent state and prevent premature differentiation.
However, we also note that the lack of KLF17 necessity in naïve hESC establishment does not rule out a more central role in pluripotency in the human embryo. To date, there have been no systematic comparisons of the outcomes of specific gene modulation in naïve hESCs versus the human pluripotent epiblast, but evidence suggests that they would not necessarily be conserved. For instance, Nanog-null naïve mESCs, while prone to differentiation, are still functionally pluripotent, with the capability for chimaera formation (Chambers et al., 2007). In contrast, a Nanog-null mouse embryo is unable to form a functional blastocyst or continue development from the peri-implantation stage onward (Mitsui et al., 2003; Chambers et al., 2003; Messerschmidt and Kemler, 2010; Frankenberg et al., 2011). Furthermore, while knockdown of POU5F1 in hESCs causes the expected differentiation phenotype (Wang et al., 2012), even partial loss of OCT4 function in the human embryo has a much more drastic phenotype, with non-cell-autonomous effects across all three lineages at the blastocyst stage (Fogarty et al., 2017). For this reason, future investigation of the function of KLF17 in human in vivo pluripotency is an important next step.

Materials and methods

Human embryo thaw and culture conditions

Human embryos at various developmental stages that were surplus to family building desires were donated to the Francis Crick Institute for use in research projects under the UK Human Fertilisation and Embryology Authority License number R0162 and Cambridge Central Research Ethics Committee number 16/EE/0067. Informed consent was obtained from all couples that donated spare embryos following IVF treatment. Before giving consent, people donating embryos were provided with all of the necessary information about the research project and an opportunity to receive counselling. Slow-frozen blastocysts (day 5 and day 6) were thawed using the BlastThaw kit (Origio; 10542010A) following the manufacturer’s instructions. Vitrified blastocysts (day 5 and day 6) were thawed using the vitrification thaw kit (Irvine Scientific; 90137-SO) following the manufacturer’s instructions. Human embryos were cultured in pre-equilibrated Global Media (Life Global) supplemented with 5 mg/ml Life Global HSA (LifeGlobal; LGPS-605) and overlaid with mineral oil (Origio; ART-4008-5P) and incubated in Embryoscope+ time lapse incubator (Vitrolife).

Maintenance of standard hESC cultures

Human embryonic stem cells (hESCs) were routinely cultured in mTeSR1 medium (Stem Cell Technologies) on growth factor-reduced Matrigel-coated dishes (BD Biosciences) and passaged as clumps at ~1:20 ratio using ReLeSR (Stem Cell Technologies). Cells were maintained in humidified incubators at 37°C, 5% CO2. The H9 cell line was obtained under licence and SLA agreement with WiCell. The H9 cell line has been exhaustively tested include STR profiling, karyotyping, gene expression, etc. The cell lines were subjected to monthly mycoplasma testing in-house and found to be negative.
**Naïve hESC culture**

All naïve hESCs were cultured at 5% O\textsubscript{2} and 5% CO\textsubscript{2}, according to recently published protocols (Guo et al., 2017; Bredenkamp et al., 2019b) on mitotically inactivated DR4 MEFs (prepared in-house) plated at a density of 1x10\textsuperscript{6} per well of a 6-well plate 12-16hrs prior to hESC seeding. Naïve hESCs were passaged as single cells by 4 minutes treatment with Accutase (Thermo Fisher) at 37°C, at split ratios between 1:3 and 1:6, every 3 to 6 days. For culture in (t)t2IL+Gö, 10 μM ROCK inhibitor (Y-27632, Tocris Bioscience) was added overnight before and after passaging, to aid survival. In-house generated, chemically reset, naïve H9 cells were maintained in tt2IL+Gö, with 0.3 μM CHIR99021 (Guo et al., 2017), and XAV supplementation until naïve passage 5.

**Generation and culture of overexpression hESC lines**

Doxycycline-inducible overexpression of HA-tagged proteins was achieved using the Lenti-X Tet-On 3G Inducible Expression System (Clontech) following the manufacturer’s protocol, and as outlined previously (Wamaitha et al., 2015). Lentiviral packaging was achieved using 7 μg of transgene-containing plasmid and the Lenti-X Packaging Single Shot reagents. Lentiviral supernatant was harvested after 48hrs and concentrated by ultracentrifugation. To produce stably transduced cells, hESCs were plated under standard conditions and changed into fresh medium the following morning. 24hrs post-plating, 10 μl concentrated virus was added to hESCs for transduction overnight (~16hrs). hESCs were dual selected with 150 μg/ml G418 and 0.5 μg/ml puromycin 48hrs post-transduction. For induction of transgene expression, doxycycline was added to mTeSR1 medium at 1 μg/ml. For the RNA-seq experiments, KLF17-inducible hESCs were plated as normal and induction initiated after 24 hours by addition of 1 μg/ml Dox to the culture medium (mTeSR1). At ~30 hours, a day 0 (pre-induction) control sample was collected, then both induced (+Dox) and uninduced (UI) samples were collected at 24hr intervals from 48hrs (day 1 post-induction) until 144hrs (day 5 post-induction). RNA was extracted from the samples and subjected to bulk RNA sequencing.

**KLF17-driven resetting of primed to naïve hESCs**

H9 KLF17-HA inducible hESCs were pre-treated overnight with 10 μM ROCKi, then harvested from standard culture (mTeSR1 on Matrigel) by 5 minutes incubation at 37°C with Accutase, resuspended in culture medium supplemented with 10 μM ROCKi and counted. 2x10\textsuperscript{5} hESCs were plated per 6-well pre-coated in DR4 MEFs and the cells placed at 5% O\textsubscript{2} and 5% CO\textsubscript{2} for ~24hrs. The following day (day 0), medium was changed to PXGL supplemented with 1 μg/ml Dox. From day 2, medium was replenished each day with PXGL freshly supplemented with 1 μg/ml Dox. On day 5, cells were passaged by 4min incubation in Accutase and plated in PXGL with 10 μM ROCKi at a split ratio between 1:5 and 1:20, dependent upon density. Within 24hrs, the cells adopted a domed morphology with highly-refractile colony edges. Cells were passaged again on day 7 or 8 and could subsequently be maintained similarly to chemically reset cells (Guo et al., 2017), with passaging every 3-4 days at split ratios of between 1:3 and 1:6.
Design of gRNAs
Guide RNAs (gRNAs) were designed in a non-biased manner against the whole cDNA sequence using a standard design tool (Hsu et al., 2013). Two strategies were attempted to achieve functional knockout of KLF17. First, the initiating methionine was targeted to potentially disrupt the entire coding sequence, leading to complete loss of KLF17 expression. Alternatively, the functional domain was targeted to directly disrupt DNA-binding or introduce a premature termination codon, leading to production of a non-functional protein. For initial screening, gRNAs were selected on the following criteria: (i) in silico score is ≥60; (ii) identified off-target sites have ≥3 mismatches; (iii) there are no (or very low frequency, ≤0.1%) single nucleotide polymorphisms (SNPs) occurring in the target sequence; (iv) the gRNA target site falls across an annotated DNA-binding domain.

Transient nucleofection of hESCs
For cell line testing of CRISPR-Cas9 efficiency, gRNAs were individually cloned into pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene plasmid #62988) (Ran et al., 2013), using the BbsI restriction sites. Nucleofection was carried out on an Amaxa 4D-Nucleofector (Lonza) with 4 μg plasmid. 24 hours prior to nucleofection, H9 hESCs were treated with 10 μM Y-27632 (Tocris Bioscience). hESCs were harvested as single cells by Accutase treatment (5min, 37°C) and counted with an automatic cell counter (Nucleocounter NC-200, ChemoMetec). For each gRNA, 2x10⁶ cells were resuspended in 100 μl P3 Primary Cell 4D-Nucleofector X Solution and transferred to nucleocuvettes with 4 μg plasmid. Nucleofection was performed with the pre-set H9 hESC programme (CB-150), then cells resuspended in antibiotic-free mTeSR1 medium supplemented with 10 μM Y-27632 and plated across half of a 6-well plate coated with DR4 MEFs to aid attachment and survival. After 24hrs, medium was changed to mTeSR1 supplemented with 0.5 μg/ml puromycin for 48hrs. Cells were allowed to recover for 8 days prior to harvesting for DNA extraction and assessment of CRISPR-Cas9 editing efficiency by MiSeq analysis. On-target editing was assessed by next-generation sequencing on the MiSeq platform (Illumina), and editing efficiency determined by analysing the FastQ files using both the Cas-Analyzer tool from CRISPR RGEN Tools (Park et al., 2017) and the CrispRVariants package in R (Lindsay et al., 2016).

Generation of clonal knockout hESCs
H9 hESCs were first nucleofected with 4 μg pSpCas9(BB)-2A-Puro (PX459) V2.0 containing the gRNA KLF17(3_1) as described above. Following 48hrs treatment with 0.5 μg/ml puromycin, cells were allowed to recover on DR4 MEFs for ~10 days, then manually passaged as single cells following treatment with Accutase (5 minutes, 37°C) or Accumax (10 minutes, 37°C; Sigma Aldrich) at clonal density into Matrigel-coated 24-well tissue culture plates (Corning). Cells were sub-cloned once more by manual picking and single-cell dissociation into 12-well plates, then 24 clones passaged in duplicate and assessed for KLF17 mutation by on-target Sanger sequencing and MiSeq analysis.
**Immunofluorescence analysis**

Cultured cells were fixed with 4% paraformaldehyde (PFA) in PBS for 1 hr at 4°C, then permeabilised in PBS containing 0.5% Tween-20 (PBS-T(0.5%)) for 20 minutes at room temperature. Blocking was carried out for 1 hr at RT in PBS-T(0.1%) with 10% donkey serum. Primary antibodies were diluted as listed in Table 1 in blocking solution, and incubated overnight at 4°C. Cells were washed several times in PBS-T(0.1%), then incubated in secondary antibodies in blocking solution for 1 hr at RT. Following repeated washing, cells were treated with DAPI-Vectashield mounting medium (Vector Labs) at 1 in 30 in PBS-T(0.1%), prior to imaging on an Olympus IX73 (Olympus Corporation). For human embryos, fixation was performed in 4% PFA in PBS for 1 hr at 4°C, then the embryos permeabilised in PBS containing 0.5% Triton-X100 (PBS-Tx(0.5%)) for 20 minutes at RT. Blocking was performed for 1 hr at RT in PBS-Tx(0.2%) containing 10% donkey serum and 3% bovine serum albumin (BSA). Primary antibodies were diluted as listed in Table 1 in blocking solution, and incubated overnight at 4°C. Embryos were washed several times, then incubated in secondary antibodies in blocking solution for 1 hr at RT. Following repeated washing, embryos were transferred into DAPI-Vectashield mounting medium (Vector Labs) at 1 in 30 in PBS-T(0.1%) on coverslip dishes (MatTek), and imaged on a Leica SP8 inverted confocal microscope (Leica Microsystems).

**RNA isolation from hESCs and qRT-PCR**

RNA was isolated using TRI reagent (Sigma) and DNase I-treated (Ambion). cDNA was synthesised using a Maxima first strand cDNA synthesis kit (Fermentas). qRT-PCR was performed using SensiMix SYBR low-ROX kit (Bioline) on a QuantStudio5 machine (Thermo Fisher). Primers pairs used are listed in Table 2. Each sample was run in triplicate. Gene expression was normalised using GAPDH as the housekeeping gene, and the results analysed using the ΔΔCt method.

**RNA sequencing**

For RNA-seq, RNA was isolated and DNase-treated as above, and libraries were prepared using the polyA KAPA mRNA HyperPrep Kit (Roche). Quality of submitted RNA samples and the resulting cDNA libraries was determined by ScreenTape Assay on a 4200 TapeStation (Agilent). Prepared libraries were submitted for single-ended 75 bp sequencing on an Illumina HiSeq 4000 (Illumina).

**Genomic DNA extraction**

Total genomic DNA was extracted from hESCs using the DNeasy Blood and Tissue Kit (Qiagen) following manufacturer’s instructions. The concentration and purity of extracted DNA was measured using a nanodrop (DeNovix).

**Protein extraction and quantification**

hESCs were harvested for protein extraction by addition of CelLytic M lysis buffer (Merck), freshly supplemented with protease inhibitors (PIC, cOmplete, EDTA-free protease inhibitor cocktail, Roche) and phosphatase inhibitors (PhlC, phosSTOP phosphatase inhibitor, Roche), directly onto plated cells. Cells were scraped, then incubated in lysis buffer for 15 min at 4°C. The lysate was collected and
clarified by centrifugation at 20,000xg for 15 min at 4°C. Protein concentration in the lysates was determined using the BCA assay, then proteins denatured by addition of 4x Laemmli sample buffer (Thermo Fisher) and heating at 90°C for 5 minutes.

Protein detection by western blotting
Denatured proteins were thawed at 65°C for 5 minutes and vortexed to ensure homogeneity. 20 μg protein per lane was loaded onto 10% Mini-PROTEAN TGX Stain-free protein gels (BIORAD), alongside 5 μl PageRuler Prestained Protein Ladder (Thermo Scientific), and electrophoresed at 100-200 V for one to two hours in a Mini-PROTEAN Tetra Vertical Electrophoresis Cell (BIORAD). Proteins were transferred onto PVDF membranes (TransBlot Turbo Mini PVDF Transfer Packs, BIORAD) using a Trans-Blot Turbo Transfer System (BIORAD). PVDF membranes were blocked for one hour in TBS-T(0.1%) containing 5% non-fat milk and incubated with primary antibodies diluted in either 5% milk or 5% BSA in TBST-T(0.1%) as shown in Table 4 overnight at 4°C. Following washes with TBS-T(0.1%), membranes were incubated with secondary antibodies in 5% milk for one hour at room temperature. Proteins of interest were visualised using the SuperSignal West Dura Extended Duration Substrate or SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) and imaged on an Amersham Imager 600RGB (GE Healthcare).

Bulk RNA-sequencing analysis
The ‘Trim Galore!’ utility version 0.4.2 was used to remove sequencing adaptors and to quality trim individual reads with the q-parameter set to 20 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (retrieved 03-05-2017)). Then sequencing reads were aligned to the human genome and transcriptome (Ensembl GRCh38 release-89) using RSEM version 1.3.0 (Li and Dewey, 2011) in conjunction with the STAR aligner version 2.5.2 (Dobin et al., 2013). Sequencing quality of individual samples was assessed using FASTQC version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (retrieved 03-05-2017)) and RNA-SeQC version 1.1.8 (DeLuca et al., 2012). Differential gene expression was determined using the R-Bioconductor package DESeq2 version 1.24.0 (Love et al., 2014). Within the DESeq2 package, adjusted p values for log-fold changes were calculated using the Benjamini-Hochberg method and the betaPrior parameter was set to “TRUE”. For the KLF17⁻/⁻ hESCs in naïve conditions, each timepoint was normalised individually, to account for the significant cell-state changes occurring across the extended time course of the experiment (~60 days). Enrichment analysis was performed using the online EnrichR tool (https://maayanlab.cloud/Enrichr/) (Chen et al., 2013; Kuleshov et al., 2016). Batch corrected PCA figures were created using the ComBat-Seq method described in Zhang et al (Zhang et al., 2020) using the R-package sva version 3.32.1.
Quantification of confocal immunofluorescence data

Nuclei were identified using StarDist (Schmidt et al., 2018) and colocalisation with fluorescent signals from KLF17 (488 nm), SOX2 (594 nm) and NANOG (647 nm) was quantified using a custom CellProfiler pipeline. Briefly, multi-channel confocal imaging Z-stacks were split into single channel image slices for preprocessing using FIJI (Schindelin et al., 2012). Nuclei were identified in image slices using the Versatile model in the StarDist 2D plugin in FIJI (Normalized Image, Percentile Low and Percentile High, 3 and 99.2% respectively, Probability Threshold and Overlap Threshold 0.5 and 0.4, respectively), using the DAPI channel as input. The output of the StarDist plugin, which consisted of images in which each segmented nucleus is assigned a unique integer pixel value in each image slice, was saved as tif files. The remaining fluorescent channels (488, 594, 647 nm) were processed with a 2-pixel radius median filter. The StarDist output image set was imported into CellProfiler v4.1.3 (McQuin et al., 2018) and the nuclei contained therein were converted into CellProfiler “Objects” using the ConvertImageToObjects module, preserving original labels. Nuclei objects were tracked through the Z-stack using Center of Mass distance-based TrackObjects module, such that an individual nucleus could be tracked through the entire Z-stack and given a unique identifier. Before tracking, nuclei objects were filtered to retain only objects bigger than 750 pixels, to create spacing between nuclei that overlap on the Z-axis. Fluorescent images were also imported into CellProfiler, background corrected and the fluorescent signal inside each nucleus was measured using the MeasureObjectIntensity module. The identifier provided by tracking the nuclei was used to aggregate the signal across all image slices for each individual nucleus using a custom MATLAB script. The segmentation output was manually checked and corrected as needed, to avoid false positive and false negative errors. All code for this project can be obtained at https://github.com/todd-fallesen/Niakan_Lab_KLF17.

Data availability

All RNA-seq data presented in this paper have been deposited on ArrayExpress, with the following accession numbers: KLF17 overexpression in primed hESCs, E-MTAB-10958; KLF17 overexpression and naïve hESCs in PXGL, E-MTAB-10915; WT and KLF17−/− hESC resetting, E-MTAB-10914.

Acknowledgements

We thank the donors of human embryos whose contributions enable this research. Further, we thank all members of the Niakan lab for their technical assistance and for help and comments on the paper, and James Briscoe, Robin Lovell-Badge and Jonathan Chubb for advice throughout the project. For additional bioinformatics support, we thank Gregorio Alanis-Lobato and Jasmin Zohren. We are grateful to the Francis Crick Institute’s Science Technology Platforms: Lyn Healy, Liani Devito and Elspeth Gamble from the Human Embryo and Stem Cell Unit; Robert Goldstone and Deb Jackson from the Advanced Sequencing Facility; Advanced Light Microscopy; and the Genomics Equipment Park.
Competing Interests
No competing interests declared.

Author Contributions
R.A.L. performed the experiments with assistance from A.M.; S.B. performed major bioinformatics analysis with assistance from R.A.L.; T.F. generated and performed the pipeline for quantification of confocal imaging; K.K.N. and R.A.L. conceived the study and analysed data; K.K.N. supervised the project; K.E., P.S., L.C., S.A., V.S., M.T. coordinated the donation of embryos to the research project; R.A.L. wrote the manuscript with feedback from all authors.

Funding
For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. Work in the laboratory of K.K.N. was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001120), the UK Medical Research Council (FC001120), and the Wellcome Trust (FC001120).

References


Heo, I., Joo, C., Cho, J., Ha, M., Han, J. & Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell, 32, 276-84.


Figures

A

KLF17  SOX2  NANOG  DAPI merge

Early day 5

Late day 5

Early day 6

Late day 6

Early day 7

B

C

Initiation of KLF17 restriction

Initiation of SOX2 restriction

% KLF17-positive nuclei co-stained with

% SOX2-positive nuclei co-stained with

Days post-fertilisation (dpf)

eD5  ID5  eD6  ID6  eD7

Days post-fertilisation (dpf)

eD5  ID5  eD6  ID6  eD7
Figure 1 – KLF17 expression in the human embryo is coincident with known pluripotency factors. (A) Representative images of immunofluorescence analysis of blastocyst stage human embryos at early day 5 (N = 5), late day 5 (N = 7), early day 6 (N = 9), late day 6 (N = 4) and early day 7 (N = 5) post-fertilisation. Scale bars = 50 μm. (B-C) The proportion of (B) KLF17-positive nuclei per embryo that are SOX2-positive or NANOG-positive and (C) SOX2-positive nuclei per embryo that are KLF17-positive or NANOG-positive. Bars represent the mean, error bars the SEM and points the percentage overlap in individual embryos.
Figure 2 – Exogenous KLF17 overexpression induces naïve factor expression in conventional hESCs. qRT-PCR analysis of H9 KLF17-HA inducible hESCs following 5 days with (+Dox) or without (UI) Dox induction of exogenous KLF17. Relative expression is displayed as fold change versus uninduced cells and normalised to GAPDH as a housekeeping gene using the ΔΔCt method. Individual samples are shown as dots, lines represent the mean and whiskers the SEM. Welch’s t test, *** $p < 0.005$, * $p < 0.05$, n.s. Not Significant, N = 3.
**Figure 3 – Exogenous KLF17 overexpression induces widespread transcriptional change in conventional hESCs.** (A) Dimensionality reduction by principal component analysis (PCA) of bulk RNA-seq data collected across a 5-day time course of H9 KLF17-HA inducible hESC growth with (+Dox) or without (UI) Dox induction of exogenous KLF17 expression. (B) Volcano plot displaying relative expression of all detected genes in +Dox versus UI H9 KLF17-HA hESCs at day 5 (logFC(+Dox Day5 vs UI Day5)) against the significance of differential expression (-log10(padj)). The red dotted line notes $p_{adj} = 0.05$. Individual genes of interest are displayed as filled circles and labelled with the gene name. (C-D) Normalised expression (transcripts per million, TPM) of individual genes of interest across the 5-day time course in UI and KLF17-expressing (+Dox) H9 KLF17-HA hESCs, showing genes that are significantly upregulated (C) or downregulated (D) at day 5. (E) Batch-corrected PCA analysis of the data from (A) integrated with published bulk RNA-seq data of samples collected before (“H9 primed”), during (“H9 Nmin”, “H9 N4+”) and following (“H9 p5 naïve”, “H9 p10 naïve”, “H9 naïve”) NANOG and KLF2 (NK2)-driven resetting of H9 hESCs (Collier et al., 2017) (F) Heatmap grouped by sample (UI or +Dox) and time point showing the genes that are highly correlated with KLF17 across time (Pearson correlation coefficient ($r$) ≥ 0.85) and that fall under the Kyoto Encyclopaedia of Genes and Genomes (KEGG) category “PI3K-Akt signalling pathway”.
Figure 4 – Exogenous KLF17 overexpression is sufficient to drive conventional hESCs to a naïve pluripotent state under PXGL culture conditions. (A) Immunofluorescence analysis H9 of KLF17-HA inducible hESCs following 5 days uninduced (UI) or 5 days doxycycline induction (+Dox) in the indicated media. Cells
were cultured on a mouse embryonic fibroblast (MEF) feeder layer and at 5% O₂. Scale bars = 20 μm. N ≥ 3. (B) Cells induced for 5 days in PXGL medium were uniquely able to give rise to typical naïve hESC-like colonies following serial bulk passaging. (C) Representative immunofluorescence analysis of H9 KLF17-HA induced naïve hESCs after 4 or 5 passages in PXGL medium. Scale bars = 20 μm. N ≥ 3. (D) Batch-corrected PCA analysis of bulk RNA-seq data tracking the progress of KLF17-driven resetting (this study) and NK2-driven resetting (Collier et al., 2017). (E) Volcano plot displaying relative expression of all detected genes in +Dox versus UI H9 KLF17-HA hESCs at day 5 (logFC(+Dox Day5 vs UI Day5)) of culture in PXGL against the significance of differential expression (-log10(padjust)). The red dotted line indicates p_adj = 0.05. Individual genes of interest are displayed as filled circles and labelled with the gene name.
Figure 5 – Generating KLF17-null mutant hESCs by CRISPR-Cas9. (A) Schematic representation of the human KLF17 locus on Chromosome 1, showing the relative position of the DNA-binding zinc finger domains (filled black rectangles) and the guide RNAs (gRNAs) tested for mutagenic efficiency. Exons are shown as red rectangles, 3’ and 5’ UTR are unfilled rectangles and introns are black chevrons. (B) Relative efficiency of each guide shown in (A) measured as a proportion of overall reads containing indel mutations following on-target amplification by MiSeq of the KLF17 target site. Dots represent individual harvested wells of CRISPR-targeted H9 hESCs, lines represent the mean and whiskers the SEM. (C) Pie charts representing the relative proportions of different outcomes of CRISPR-Cas9 editing.
of H9 hESCs, based on the sequences detected by MiSeq analysis. (D) Immunofluorescence analysis of H9 hESCs targeted with Cas9 and gRNA KLF17(3_1) by, following epigenetic resetting (Guo et al., 2017) for 8 days. Internal wild-type (WT) controls (#7, #15 and #21) are clones that were subjected to nucleofection, puromycin selection and clonal expansion, but were unedited, with a wild-type genotype. Compound null mutant clones (#18 and #19) were verified by MiSeq and IF. N = 3.
Figure 6 – *KLF17*-null hESCs are capable of attaining and maintaining naïve pluripotency. (A) Representative brightfield images of WT and *KLF17*-/- H9 hESCs following 10 passages under naïve culture conditions. (B) Dimensionality reduction by principal component analysis (PCA) of bulk RNA-seq data collected at various times during the epigenetic resetting (Guo et al., 2017) of WT and *KLF17*-/- H9 hESCs. (C-F) Normalised expression (TPM) of individual genes of interest across the full resetting time course showing (C) lack of appreciable *KLF17* transcripts, (D) temporally limited upregulation of the parologue *KLF5* in *KLF17*-/- H9 hESCs and equivalent expression of the paralogues (E) *KLF2* and (F) *KLF4*. (G) Volcano plot displaying relative expression of all detected genes in *KLF17*-/- versus WT naïve H9 hESCs following 5 passages in naïve culture conditions (logFC(*KLF17*-/- p5 vs WT p5)) against the significance of differential expression (-log10(padj)). The red dotted line indicates padj = 0.05. Individual genes of interest are displayed as filled circles and labelled with the gene name. (H-I) Normalised expression (TPM) of individual genes of interest across the full resetting time course showing (H) downregulation of the pluripotency-associated factor *LIN28A* and (I) upregulation of the WNT signalling receptor *FZD5* in *KLF17*-/- H9 hESCs.
### Table 1 – Primary and secondary antibodies used in immunofluorescence

<table>
<thead>
<tr>
<th>Target</th>
<th>Species</th>
<th>Dilution</th>
<th>Supplier</th>
<th>Catalogue Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-DNMT3L</td>
<td>Mouse</td>
<td>1 in 500</td>
<td>Abcam</td>
<td>ab93613</td>
</tr>
<tr>
<td>Anti-DPPA5</td>
<td>Rabbit</td>
<td>1 in 250</td>
<td>Sigma Aldrich</td>
<td>D2569</td>
</tr>
<tr>
<td>Anti-GP130</td>
<td>Rabbit</td>
<td>1 in 250</td>
<td>Thermo Fisher</td>
<td>PA5-80735</td>
</tr>
<tr>
<td>Anti-HA (3F10)</td>
<td>Rat</td>
<td>1 in 500</td>
<td>Sigma Aldrich (Roche)</td>
<td>11867423001</td>
</tr>
<tr>
<td>Anti-KLF5</td>
<td>Rabbit</td>
<td>1 in 500</td>
<td>Abcam</td>
<td>ab137676</td>
</tr>
<tr>
<td>Anti-KLF17</td>
<td>Rabbit</td>
<td>1 in 500 (hESCs)</td>
<td>Atlas Antibodies</td>
<td>HPA024629</td>
</tr>
<tr>
<td>Anti-NANOG</td>
<td>Goat</td>
<td>1 in 200</td>
<td>R&amp;D Systems</td>
<td>AF1997</td>
</tr>
<tr>
<td>Anti-OCT4</td>
<td>Mouse</td>
<td>1 in 100</td>
<td>Santa Cruz Biotechnology</td>
<td>SC-5279</td>
</tr>
<tr>
<td>Anti-SOX2</td>
<td>Rat</td>
<td>1 in 100</td>
<td>Invitrogen</td>
<td>14-9811-82</td>
</tr>
<tr>
<td>Anti-SUSD2</td>
<td>Mouse</td>
<td>1 in 250</td>
<td>Biolegend</td>
<td>327401</td>
</tr>
<tr>
<td>Anti-TFAP2C</td>
<td>Goat</td>
<td>1 in 200</td>
<td>R&amp;D Systems</td>
<td>AF5059</td>
</tr>
<tr>
<td>Anti-VENTX</td>
<td>Rabbit</td>
<td>1 in 500</td>
<td>Cambridge Bioscience</td>
<td>HPA050955</td>
</tr>
<tr>
<td>Alexa Fluor anti-mouse IgG</td>
<td>Donkey</td>
<td>1 in 300</td>
<td>Invitrogen</td>
<td>A21202 (488 nm) A21203 (594 nm) A31571 (647 nm)</td>
</tr>
<tr>
<td>Alexa Fluor anti-rabbit IgG</td>
<td>Donkey</td>
<td>1 in 300</td>
<td>Invitrogen</td>
<td>A21206 (488 nm) A21207 (594 nm) A31573 (647 nm)</td>
</tr>
<tr>
<td>Alexa Fluor anti-goat IgG</td>
<td>Donkey</td>
<td>1 in 300</td>
<td>Invitrogen</td>
<td>A11055 (488 nm) A11058 (594 nm) A21447 (647 nm)</td>
</tr>
<tr>
<td>Alexa Fluor anti-rat IgG</td>
<td>Donkey</td>
<td>1 in 300</td>
<td>Invitrogen</td>
<td>A21208 (488 nm) A21209 (594 nm)</td>
</tr>
</tbody>
</table>
Table 2 – Primers used for qRT-PCR

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer sequence</th>
<th>Reverse primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARGFX</td>
<td>CCAGTTTCACTCTGTATCCAAG</td>
<td>CGTTCTTTATGCCTTCTCCG</td>
</tr>
<tr>
<td>DNMT3L</td>
<td>GGACCCCTCGATCTTGTA</td>
<td>ACCAGATTTGCCACGAACAT</td>
</tr>
<tr>
<td>DPPA5</td>
<td>GTGGTTTACGGCTCCTATT</td>
<td>TCATCCAAGGGCCTAGTT</td>
</tr>
<tr>
<td>GAPDH</td>
<td>GATGACATCAAGAAGGGTGTG</td>
<td>GTCTACATGGCAACTGTGAGG</td>
</tr>
<tr>
<td>KLF17</td>
<td>ACCCAGTCTTCATGTACGCC</td>
<td>GCACCTCCAGAGCTCCAGAA</td>
</tr>
<tr>
<td>KLF17_HA</td>
<td>ACACCAGAAGACTCATCGGC</td>
<td>ACATCGTATGGGTAAGGACCAG</td>
</tr>
<tr>
<td>NANOG</td>
<td>CATGAGTGTTGATCCAGCTTG</td>
<td>CCTGAATAAGCAGATCCATGG</td>
</tr>
<tr>
<td>ZFP42</td>
<td>GGAATGTGGAGAAGCGTTCGT</td>
<td>CGGTGGTCGTCGCAGCT</td>
</tr>
<tr>
<td>TFCP2L1</td>
<td>AGCACATCCACCGAGTCTAC</td>
<td>TGAGGACAAAAACAGGATTTCT</td>
</tr>
<tr>
<td>VENTX</td>
<td>CAGCTTCAAGAGTGGAGTAAGA</td>
<td>AGACGTGAGTAGAAGCTGG</td>
</tr>
</tbody>
</table>

Table 3 – Primers used for genotyping the KLF17 on-target locus

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer sequence</th>
<th>Reverse primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1</td>
<td>GTGGCGATGTACCGATAACC</td>
<td>CTCCGCCTCACCTCTCTCT</td>
</tr>
<tr>
<td>Exon 3</td>
<td>GGACCTTCCTTTGGAATCCTC</td>
<td>CGGCTGCAGCTGACACATTT</td>
</tr>
<tr>
<td>Long-range</td>
<td>ACAGGTTGGAAGGTGTCAG</td>
<td>GAAGCTGGTCAAGGCAGGTA</td>
</tr>
<tr>
<td>Target</td>
<td>Species</td>
<td>Dilution</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>Anti-alpha tubulin</td>
<td>Mouse</td>
<td>1 in 1000</td>
</tr>
<tr>
<td>Anti-pan AKT</td>
<td>Mouse</td>
<td>1 in 2000</td>
</tr>
<tr>
<td>Anti-phospho AKT (Ser473)</td>
<td>Rabbit</td>
<td>1 in 2000</td>
</tr>
<tr>
<td>Anti-phospho AKT (Thr308)</td>
<td>Rabbit</td>
<td>1 in 1000</td>
</tr>
<tr>
<td>Anti-pan ERK1/2</td>
<td>Mouse</td>
<td>1 in 2000</td>
</tr>
<tr>
<td>Anti-phospho ERK1/2</td>
<td>Rabbit</td>
<td>1 in 2000</td>
</tr>
<tr>
<td>Anti-pan IGF1R</td>
<td>Rabbit</td>
<td>1 in 1000</td>
</tr>
<tr>
<td>Anti-pan Insulin Receptor (InsR)</td>
<td>Mouse</td>
<td>1 in 1000</td>
</tr>
<tr>
<td>Anti-phospho IGF1R/InsR</td>
<td>Rabbit</td>
<td>1 in 1000</td>
</tr>
<tr>
<td>Anti-KLF17</td>
<td>Rabbit</td>
<td>1 in 500</td>
</tr>
<tr>
<td>Anti-pan S6</td>
<td>Mouse</td>
<td>1 in 1000</td>
</tr>
<tr>
<td>Anti-phospho S6</td>
<td>Rabbit</td>
<td>1 in 1000</td>
</tr>
<tr>
<td>Anti-mouse IgG (H+L), HRP-conjugated</td>
<td>Goat</td>
<td>1 in 20000</td>
</tr>
<tr>
<td>Anti-rabbit IgG (H+L), HRP-conjugated</td>
<td>Goat</td>
<td>1 in 20000</td>
</tr>
<tr>
<td>Anti-goat IgG (H+L), HRP-conjugated</td>
<td>Donkey</td>
<td>1 in 20000</td>
</tr>
<tr>
<td>Anti-rat IgG (H+L), HRP-conjugated</td>
<td>Goat</td>
<td>1 in 20000</td>
</tr>
</tbody>
</table>
Fig. S1. – KLF17 expression in the human embryo is coincident with known pluripotency factors. (A) Immunofluorescence analysis of blastocyst stage human embryos at early day 5 (N = 5), late day 5 (N = 7), early day 6 (N = 9), late day 6 (N = 4) and early day 7 (N = 5) post-fertilisation. Scale bars = 50 μm. (B) Quantification of the number of segmented DAPI-positive nuclei per embryo across blastocyst development. (C) Quantification of the total proportion of DAPI-positive nuclei per embryo that are KLF17-positive, SOX2-positive or NANOG-positive. Bars represent the mean, error bars the SEM and points the percentage in individual embryos.
Fig. S2. – Generating primed hESCs to inducibly express ectopic HA-tagged KLF17. (A) Schematic diagram showing the generation of H9 KLF17-HA inducible hESCs via lentiviral transduction. (B) Immunofluorescence analysis of H9 KLF17-HA inducible hESCs following 5 days uninduced (UI) or 5 days doxycycline induction (+Dox). Scale bars = 20 μm. N ≥ 3.
Fig. S3. – 70 genes are strongly correlated with KLF17 expression over time. (A) Heatmap ordered by sample (UI or +Dox) and time point showing all genes that are highly correlated with KLF17 across time (Pearson correlation coefficient (r) ≥ 0.85). (B-C) Volcano plots displaying relative expression of all detected genes in +Dox versus UI H9 KLF17-HA hESCs at (B) day 1 (logFC(+Dox Day1 vs UI Day1)) and (C) day 2 (logFC(+Dox Day2 vs UI Day2)) against the significance of differential expression (-log10(padj)). The red dotted line indicates padj = 0.05. All genes with correlation coefficient to KLF17 ≥ 0.85 are displayed as filled circles and genes associated with PI3K-AKT signalling (as shown in Fig. 3F) are labelled with the gene name.
Fig. S4. – Genes highly correlated with KLF17 expression include numerous signalling components and cytoskeletal/ECM components. (A-P) Normalised expression (TPM) of individual genes of interest across the 5-day time course showing factors involved in (A-G) PI3K-AKT signalling, (H-I) TGFβ signalling, (J-L) other signalling pathways, (F-G,M-O) the cytoskeleton/ICM and (P) the pluripotency-regulating long non-coding RNA LINC-ROR. (Q) A heatmap depicting the relative expression, based on TPM values, of the genes of interest from A-P and Fig. 3C-D in U1 and +Dox H9 KLF17-HA hESCs alongside established primed and naive hESCs (Collier et al., 2017).
A. Day 1, 2, 5

<table>
<thead>
<tr>
<th></th>
<th>Dox</th>
<th>-</th>
<th>+</th>
<th>-</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>55 kDa</td>
<td>KLF17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 kDa</td>
<td>pIGF1R/InsR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 kDa</td>
<td>IGF1R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 kDa</td>
<td>InsR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 kDa</td>
<td>pAKT_Ser473</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 kDa</td>
<td>pAKT_Thr308</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 kDa</td>
<td>AKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 kDa</td>
<td>pS6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 kDa</td>
<td>S6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 kDa</td>
<td>pERK1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40 kDa</td>
<td>ERK1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 kDa</td>
<td>α-Tubulin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. KLF17

Expression level (U/L) vs. Day

C. pAKT-Ser473

Fold change (U/L = 1) vs. Day

D. pAKT-Thr308

Fold change (U/L = 1) vs. Day

E. pS6

Fold change (U/L = 1) vs. Day

F. pIGF1R/InsR

Fold change (U/L = 1) vs. Day

G. pERK1/2

Fold change (U/L = 1) vs. Day
Fig. S5. – KLF17 overexpression brings about changes in PI3K-AKT signalling pathway activity. (A) Representative western blot analysis of ectopic KLF17 induction following 1, 2 and 5 days Dox treatment of H9 KLF17-HA and the associated changes in phosphorylation of various components of the PI3K-AKT signalling pathway. (B-G) Quantification of the levels of protein detection by western blot showing (B) the steady increase in KLF17 protein levels (arbitrary units, AU) in +Dox versus UI H9 KLF17-HA hESCs and (C-G) the dynamic changes in phosphorylation and activation status of various components of the PI3K-AKT signalling pathway, represented as a fold-change of +Dox versus UI. All intensity values were normalised to the level of the α-Tubulin loading control. N = 3.
Fig. S6. — Further investigation of putative KLF17 target genes. (A-E) Example tracks from the IGV browser, showing the binding sites of KLF17 near to selected genes of interest, as identified in naive hESCs cultured under HENSM conditions (Bayerl et al., 2021). (F-K) qRT-PCR analysis across the 5-day time course of Dox treatment in H9 KLF17-HA hESCs. Relative expression is displayed as fold change versus uninduced cells and normalised to GAPDH as a housekeeping gene using the ΔΔCt method. Dots represent the mean and whiskers the SEM. Welch’s t test, **** p < 0.001, *** p < 0.005, ** p < 0.01, * p < 0.05, n.s. Not Significant, N = 3. (L) Immunofluorescence analysis of H9 KLF17-HA inducible hESCs following 5 days uninduced (UI) or 5 days doxycycline induction (+Dox). Scale bars = 20 μm. N ≥ 3.
Fig. S7. – PXGL is uniquely able to support KLF17-driven naïve resetting of H9 KLF17-HA hESCs (A) Immunofluorescence analysis of H9 KLF17-HA inducible hESCs following 5 days uninduced (UI) or 5 days doxycycline induction (+Dox) in the indicated media. Cells were cultured on a mouse embryonic fibroblast (MEF) feeder layer and at 5% O₂. Scale bars = 20 μm. N ≥ 3. (B) Unlike induced cells, UI control H9 KLF17-HA cells were unable to survive in PXGL medium following the first passage. Under mTeSR1 or t2iL+Go conditions, neither +Dox nor UI cells transitioned to naïve morphology.
Fig. S8. – KLF17 expression in PXGL drives upregulation of naïve-associated genes and downregulation of primed-associated genes. Normalised expression (TPM) of individual genes of interest across the 5-day time course of H9 KLF17-HA in PXGL. (A) Examples of naïve hESC-associated genes whose expression is induced only in the presence of ectopic KLF17 expression. (B) Examples of primed hESC-associated genes whose expression is specifically restrained when KLF17 expression is induced. (C) Examples of genes whose expression dynamics are similar irrespective of KLF17 expression.
Fig. S9. — Generating KLF17-null hESCs by CRISPR-Cas9. (A) A table showing the gRNA sequences tested for mutagenic efficiency. (B) Pie charts representing the relative proportions of different outcomes of CRISPR-Cas9 editing of H9 hESCs, based on the sequences detected by MiSeq analysis.
A.

<table>
<thead>
<tr>
<th>Clone #</th>
<th>1</th>
<th>4</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>15</th>
<th>18</th>
<th>19</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-range genotype</td>
<td>Δ8</td>
<td>Δ16/Δ21</td>
<td>WT</td>
<td>Δ2</td>
<td>Δ9</td>
<td>Δ8</td>
<td>Δ8</td>
<td>WT</td>
<td>+1/Δ8</td>
<td>+1/Δ8</td>
<td>WT</td>
</tr>
<tr>
<td>Long-range genotype</td>
<td>Δ8/7</td>
<td>Δ16/Δ21</td>
<td>WT</td>
<td>Δ2/7</td>
<td>Δ9/Δ163</td>
<td>Δ8/7</td>
<td>WT</td>
<td>+1/Δ8</td>
<td>+1/Δ8</td>
<td>WT</td>
<td></td>
</tr>
</tbody>
</table>

B.

```
393 bp
---
262 bp
---
123 bp
---

rs35084281 (C/C)  KLF17(3_1) cut site
rs34057178 (C/A)  180 bp
```

C.

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Parental H9</th>
<th>#7</th>
<th>#1</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
<th>#11</th>
</tr>
</thead>
<tbody>
<tr>
<td>C / CAA</td>
<td>53.85%</td>
<td>37.5%</td>
<td>0%</td>
<td>0%</td>
<td>29.17%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>- / -</td>
<td>46.15%</td>
<td>62.5%</td>
<td>100%</td>
<td>100%</td>
<td>70.83%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

**D.**

**Zinc finger domains**

<table>
<thead>
<tr>
<th>Reference target region</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>KLF17(3_1) PAM</td>
<td></td>
</tr>
</tbody>
</table>

**E.**

**Δ8 bp deletion**

**F.**

**+1 bp insertion**

**HxNH motif**

**Premature termination codons - third zinc finger is lost**

**Premature termination codons - third zinc finger is lost**

**HxNH motif is disrupted**

**HxNH motif is disrupted**
Fig. S10. -- Genotyping of H9 hESCs following targeting with gRNA KLF17(3_1) and clonal expansion. (A) A table showing the results of genotyping 11 clones generated following CRISPR-Cas9 targeting. Short-range genotype denotes the results of MiSeq of a ~250 bp region surrounding the KLF17(3_1) cut site. Long-range genotype denotes the results of Sanger sequencing of a ~950 bp region surrounding the KLF17(3_1) cut site. The red rectangle highlights the verified KLF17-c H9 hESCs that were carried forward. (B) Schematic of the short- and long-range genotyping approach employed on the 11 clones in (A). (C) A table showing the percentage of interpretable reads that showed one of two possible variant-types at the highly polymorphic regions illustrated in (B) -- rs35084281 and rs34057178. Parental H9 is the unmodified control cell line, #7 is an internal wild-type control clone generated following nucleofection of KLF17(3_1), #1, #6, #9, 10 and #11 are the KLF17-targeted H9 clones that appeared to have undergone homozygous editing based on short-range genotyping. (D-F) Illustration of the sequence context surrounding the KLF17(3_1) cut site in (D) the wild-type reference sequence, (E) the case of an 8 bp deletion and (F) the case of a 1 bp insertion. Important features of the KLF17 sequence are highlighted. DNA sequence is shown in regular font, amino acid sequence is bold above or below the DNA.
Fig. S11. – *KLF17*-null naive hESCs at passage 5 display misregulated expression of core glycolytic enzymes and WNT pathway components. (A) Schematic representation of the chemical epigenetic resetting experiment, showing the timings of mRNA collections. (B) Batch-corrected PCA analysis of the bulk RNA-seq data shown in Fig. 6B incorporated with data from NK2-driven resetting (Collier et al., 2017). (C–J) Normalised expression (TPM) of individual genes of interest across the resetting protocol showing (C–H) downregulation of glycolytic enzymes and (I–J) upregulation of WNT signalling factors at p5.
Table S1. - Genes significantly differentially expressed following 5 days expression of ectopic KLF17 in primed hESCs

Click here to download Table S1

Table S2. - Genes whose expression is highly correlated (r ≥ 0.85) to ectopic KLF17

Click here to download Table S2

Table S3. - Genes associated with WNT signalling and significantly downregulated following 24hrs ectopic KLF17 expression in primed hESCs

Click here to download Table S3

Table S4. - Genes significantly differentially expressed following 5 days expression of ectopic KLF17 in hESCs cultured in PXGL medium

Click here to download Table S4

Table S5. - Genes significantly differentially expressed following 5 passages in naive conditions of KLF17⁺ hESCs

Click here to download Table S5

Table S6. - Genes associated with WNT signalling and significantly upregulated following 5 passages in naive conditions of KLF17⁺ hESCs

Click here to download Table S6