The transcriptional regulator ZNF398 mediates pluripotency and epithelial character downstream of TGF-beta in human PSCs

Irene Zorzan, Marco Pellegrini, Mattia Arboit, Danny Incarnato, Mara Maldotti, Mattia Forcato, Guidantonio Malagoli Tagliazucchi, Elena Carbognin, Marco Montagner, Salvatore Oliviero & Graziano Martello

Human pluripotent stem cells (hPSCs) have the capacity to give rise to all differentiated cells of the adult. TGF-beta is used routinely for expansion of conventional hPSCs as flat epithelial colonies expressing the transcription factors POU5F1/OCT4, NANOG, SOX2. Here we report a global analysis of the transcriptional programme controlled by TGF-beta followed by an unbiased gain-of-function screening in multiple hPSC lines to identify factors mediating TGF-beta activity. We identify a quartet of transcriptional regulators promoting hPSC self-renewal including ZNF398, a human-specific mediator of pluripotency and epithelial character in hPSCs. Mechanistically, ZNF398 binds active promoters and enhancers together with SMAD3 and the histone acetyltransferase EP300, enabling transcription of TGF-beta targets. In the context of somatic cell reprogramming, inhibition of ZNF398 abolishes activation of pluripotency and epithelial genes and colony formation. Our findings have clear implications for the generation of bona fide hPSCs for regenerative medicine.
Human pluripotent stem cells (hPSCs) have been derived from human blastocysts as human embryonic stem cells (hESCs) or from somatic cells via transcription factor-mediated reprogramming as induced pluripotent stem cells (hiPSCs)\(^1\). Initially hPSCs were cultured on layers of inactivated fibroblasts (feeder cells), which produce several adhesion and signalling molecules. Alternatively, the medium was conditioned, or enriched by unknown secreted factors, by fibroblasts before their use with hPSCs. Such poorly defined culture systems represented a hurdle to the identification of key signals regulating pluripotency. Importantly, chemically defined conditions for the expansion of hPSCs have been reported\(^2\)–\(^7\). Despite variations in the media composition, ligands of the TGF-beta family are invariably added or produced by feeder cells\(^8\),\(^9\). Indeed, TGF-beta signalling has been shown to be critical for the maintenance of pluripotency in hPSCs\(^10\),\(^11\). However, the mechanisms of action of the TGF-beta signal remain poorly characterised.

TGF-beta ligands such as TGF-beta1/2/3 (TGFBI/2/3), Nodal and Activin A bind a dimer of type II serine/threonine kinase receptors, which in turn phosphorylate and activate two type I receptors, leading to the formation of a hetero-tetrameric receptor complex. Activation of the receptor complex leads to phosphorylation of SMAD2 and SMAD3, the receptor-SMADs (R-SMADs). Phosphorylated R-SMADs form heteromeric complexes with SMAD4 and translocate into the nucleus, where they bind target genes.

SMAD3 binds the DNA directly, while SMAD2 needs SMAD4 to do so\(^12\),\(^13\), in combination with the histone acetyltransferase EP300, ultimately leading to activation of target genes. R-SMADs are known to interact with additional transcription factors that may vary between different cell types\(^15\), resulting in activation of cell type-specific transcriptional programmes (see Supplementary Fig. 1a for a diagram of the TGF-beta pathway)\(^12\),\(^13\). In order to understand how TGF-beta signalling regulates the behaviour of hPSCs, it is critical to identify genes directly induced by R-SMADs.

Core pluripotency factors—POU5F1/OCT4, NANOG, SOX2—were initially identified in murine naïve pluripotent cells\(^16\),\(^17\) and were then found to be functionally relevant in hPSCs\(^19\). A large set of additional murine pluripotency factors have been identified\(^20\), the majority of which are not expressed in conventional human PSCs, potentially because of differences between species or because conventional hPSCs are in a more advanced developmental state called primed pluripotency. Although naïve hPSCs have been recently generated either directly from embryos or by reprogramming of somatic cells\(^21\)–\(^24\), they are not the focus of this study and, for clarity, we should stress that the acronyms hPSCs, hESCs and hiPSCs indicate only human conventional pluripotent cells in a primed state.

Here, we study conventional hPSCs with the aim of isolating human-specific pluripotency regulators that could reveal differences between PSCs of different species, or could play a critical role for induction of human pluripotency.

In this study, we characterise the transcriptional programme activated by TGF-beta/SMAD3 signalling in hPSCs. We identify several potential downstream mediators and test them using a gain-of-function approach. TGF-beta appears to maintain pluripotency via induction of four factors. Among them, we extensively characterise a transcriptional regulator, called ZNF398, which induces genes associated with pluripotency and epithelial character in collaboration with SMAD3 and the histone acetyltransferase EP300. Moreover, ZNF398 knockdown during somatic cell reprogramming causes a drastic reduction in iPSC colonies.

**Results**

**Identification of TGF-beta transcriptional targets in hPSCs.** We expanded hPSCs under chemically defined conditions\(^4\),\(^5\) and validated the known role of TGF-beta in maintenance of pluripotency using SB431542 (SB43), an inhibitor of TGFBR1 and ACVR1B/C, the type I receptors mediating TGFBI/2/3, Activin and Nodal signalling (Supplementary Fig. 1a). SB43 reduced phosphorylation of SMAD3 downstream of TGFBI and reduced the levels of the pluripotency factors POU5F1/OCT4, PRDM14 and NANOG (Fig. 1a and Supplementary Fig. 1b) as previously reported\(^2\)–\(^7\).

Human pluripotent colonies are composed of a monolayer of cells expressing epithelial markers. SB43 treatment induces also a morphological change, with loss of cell–cell contact, reduction of epithelial markers and upregulation of mesenchymal markers (Fig. 1b and Supplementary Fig. 1c), as previously described\(^25\),\(^26\).

We decided to study how TGF-beta controls gene programmes associated with pluripotency and the epithelial character with an unbiased functional approach based on the identification of direct transcriptional targets followed by functional validation\(^26\),\(^27\). We reasoned that TGF-beta transcriptional targets should be bounded by SMAD2/3 and either downregulated upon signal inhibition or rapidly induced upon stimulation. SMAD2 and SMAD3 can also form heterodimers\(^14\) and have redundant functions in pluripotent cells\(^28\). We focused on SMAD3, given that it is more abundant than SMAD2 in hPSCs and it binds directly to the DNA\(^12\),\(^13\) (Supplementary Fig. 1d). We intersected SMAD3 chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) with gene expression data from cells treated with SB43 and identified 195 genes downregulated and bound by SMAD3 (Fig. 1c and d, top panel yellow dots on the left). Moreover, by intersecting SMAD3-bound genes with genes induced after 4 h of acute stimulation, we identified 61 additional putative targets and 20 that were also among the downregulated genes (Fig. 1c and d, bottom panel yellow dots on the right). Several known TGF-beta direct targets, such as LEFTY1/2, SKIL and SMAD7, were identified (Fig. 1d), supporting the validity of our approach.

We then refined our gene list by focusing on genes encoding for transcriptional regulators, such as transcription factors or chromatin modifiers, given that such classes of proteins have the capacity to direct transcriptional programmes. Finally, we included only genes robustly expressed (>3 RPKM in hPSCs) (Supplementary Data 1). We performed qPCR to independently validate our putative TGF-beta targets. In particular, we tested the responsiveness of TGFBI and Activin A, two ligands commonly used for hPSCs expansion\(^4\)–\(^7\) (Supplementary Fig. 1a and f). We also tested whether targets were responsive to the TGF-beta signal when cells were expanded either on feeders or under feeder-free conditions, given that the TGF-beta signal is active and maintains pluripotency under both conditions (Supplementary Fig. 1g)\(^5\),\(^8\),\(^9\). After extensive validation, we identified eight genes (ID1, MYC, BOCR, KLF7, OTX2, ZNF398, NANOG and ETS2) as bona fide TGF-beta and Activin A transcriptional targets in hPSCs (Fig. 2a, b).

**Functional identification of pluripotency regulators.** If a gene is a critical downstream mediator of the TGF-beta signal in hPSCs, its forced expression should maintain pluripotency also when TGF-beta signalling is inhibited. To test this hypothesis, we stably expressed our candidates in hPSCs and hESCs using piggyBac (PB) vectors.

First, we performed a clonal assay, which allows us to quantify the fraction of hPSCs able to self-renew, giving rise to pluripotent colonies. Cells transfected with an empty vector formed a reduced number of alkaline phosphatase (AP)-positive pluripotent colonies when treated with SB43 (Fig. 3a and Supplementary...
Only expression of NANOG, KLF7, MYC and ZNF398 resulted in full rescue in formation of AP-positive colonies in the presence of SB43, while other factors had only a partial, or no effect. Second, under forced expression of either NANOG, KLF7, MYC or ZNF398 the cells maintained a epithelial-like morphology, generally associated with pluripotency (Fig. 3b), upon SB43 treatment, while other factors failed to do so. Third, NANOG, KLF7, MYC and ZNF398 were each able to maintain expression of pluripotency markers (Fig. 3c) in presence of SB43, although they displayed specificity for different targets. For instance, ZNF398 and NANOG activated robustly PRDM14 expression. Quantitative immunostaining confirmed maintenance of OCT4 and NANOG at the protein levels (Fig. 4a). Comparable results were also obtained after prolonged culture with SB43 (Supplementary Fig. 2b–d). We confirmed our results in an additional hESC line (Supplementary Fig. 3a–c) and confirmed a
similar level of transgene expression in different cell lines for different constructs (Supplementary Fig. 3d).

Taken together these results indicate that forced expression of NANOG, KLF7, MYC and ZNF398 is individually sufficient to stably maintain an ESC-like state upon TGF-beta inhibition. Thus, we identified from the literature an extended suite of functional regulators of human pluripotency: FOXO129, PRDM1430, BCOR31, LIN28A3, LIN28B32, DPPA2/433, SOX219, UTF134 and, in the present work, we identified the transcription factors KLF7 and ZNF398.

We noticed that MYC had a strong effect on AP-positive colony formation and morphology despite the partial effects on OCT4 and NANOG expression (Figs. 3a, b and 4a), suggesting that MYC might maintain pluripotency via other pluripotency factors.
Fig. 2 Validation of TGF-beta transcriptional targets. a Balloon plot summarising validation experiments. Balloon size indicates the statistical significance, colour indicates the fold-change in expression relative to DMSO-treated hPSCs. Left: Microarray data as in top panel of Fig. 1d were independently validated by qPCR. Validation experiments were performed in two different culture conditions: feeder-free or on feeders (MEF). Expression was normalised to the mean of DMSO-treated cells. Right: RNA-seq data as in bottom panel of Fig. 1d were independently validated by qPCR. Expression was normalised to the mean of SB43-treated samples. The grey box highlights the known direct targets of SMAD3 (positive controls, including SKIL, one of the 21 candidates). Transcriptional SMAD3 targets independently confirmed are highlighted in bold. For qPCR validation, five independent experiments were performed for each condition. Unpaired two-tailed t-test, p-values were not adjusted. See also Supplementary Fig. 1f, g. Source data are provided as a Source Data file.

b Example of SMAD3 binding and gene expression analysis of three validated targets. Top: Gene tracks represent binding of SMAD3 (data from ref.15) at the indicated gene loci. Red lines indicate the regions validated by ChIP-qPCR. Center: ChIP-qPCR on SMAD7, NANOG and ZNF398 loci was performed using anti-SMAD3 and anti-SMAD2 or a rabbit control IgG antibody in BG01V (dark grey) and H9 (light grey) cell lines. Enrichment is expressed as a percentage of the DNA inputs. Bars indicate the mean of two biological replicates shown as dots. Bottom: Gene expression analysis by qPCR. Bars indicate mean ± SEM of n = 7, 5, 7, 9, 10, 8, 6 biological replicates over six independent experiments shown as dots for SMAD7, n = 7, 6, 5, 6, 5, 5, 5 biological replicates over five independent experiments shown as dots for NANOG, and n = 8, 7, 5, 6, 9, 10, 8, 6 biological replicates over six independent experiments shown as dots for ZNF398. Expression was normalised to the mean of SB43 16 h samples. Unpaired two-tailed t-test. Source data are provided as a Source Data file.

Fig. 3 Functional identification of pluripotency regulators in hPSCs. a Left: Clonal assay quantification of hESCs (HES2, red bars) and hiPSCs (KIPS, orange bars) stably expressing an empty vector control (Empty) or eight different SMAD3 targets identified in Fig. 2a. Two thousand cells were seeded at clonal density in the presence of DMSO or SB43 and stained for alkaline phosphatase (AP) after 5 days. Bars show the mean ± SEM percentage of AP-positive colonies. Dots represent independent experiments (n = 7, 10 for Empty DMSO; n = 7, 10 for Empty SB43; n = 3, 4 for NANOG SB43; n = 2, 2 for KLF7 SB43; n = 4, 3 for MYC SB43; n = 3, 2 for ZNF398 SB43; n = 3, 2 for ID1 SB43; n = 2, 1 for BCOR SB43; n = 2, 1 for ETS2 SB43; n = 3, 2 for ID1 SB43). Bars indicate the mean ± SEM of independent experiments shown as dots for SMAD7, n = 7, 6, 5, 6, 5, 5, 5, 5 biological replicates over five independent experiments shown as dots for NANOG, and n = 8, 7, 5, 6, 9, 10, 8, 6 biological replicates over six independent experiments shown as dots for ZNF398. Expression was normalised to the mean of SB43 16 h samples. Unpaired two-tailed t-test. Source data are provided as a Source Data file. Right: Representative images of clonal assay performed in KIPS. See also Supplementary Fig. 3a for results obtained in H9 hESCs. Scale bars 500 µm. Source data are provided as a Source Data file.

b Morphology of HES2 colonies stably expressing an empty vector (Empty) in presence of DMSO or SB43 and HES2 stably expressing the eight SMAD3 targets in presence of SB43. Representative images of three independent experiments are shown. See also Supplementary Fig. 3b for results obtained in H9 hESCs. Scale bars 200 µm. c Gene expression analysis by qPCR of HES2 (light green bars) and KiPS (dark green bars) stably expressing an Empty vector or the eight SMAD3 targets and treated with or without SB43 for 5 days. Bars indicate mean ± SEM of independent experiments, shown as dots (n = 5, 5 for NANOG overexpression; n = 2, 2 for KLF7 overexpression; n = 4, 4 for MYC overexpression; n = 4, 4 for ZNF398 overexpression, n = 2, 4 for ID1 overexpression; n = 2, 2 for BCOR overexpression; n = 2, 1 for ETS2 overexpression; n = 1, 2 for OTX2 overexpression in HES2 and KIPS, respectively). Expression was normalised to the Empty DMSO samples. Unpaired two-tailed Mann–Whitney U test. Source data are provided as a Source Data file.
performed transcriptome analysis and observed that MYC robustly activated UTF1, KLF7, LIN28B and DPPA2, despite the mild effect on NANOG and OCT4 (Fig. 4b). Similarly, NANOG, KLF7 and ZNF398 activated completely different sets of pluripotency regulators. We further extended our analysis to all genes highly expressed in hPSCs that were significantly reduced upon SB43 treatment for 5 days. We identified 538 genes downregulated by SB43 treatment (5 days) in KiPS stably expressing an empty vector (Supplementary Fig. 4b) and none of the four factors were able to maintain the expression of the markers analysed, with the exception of Otx2 maintained only by KLF7. We conclude that the ability of NANOG, KLF7, MYC and ZNF398 to maintain pluripotency is not conserved in murine EpiSCs.

In sum, our results indicate that in hPSCs, TGF-beta maintains pluripotency mainly via a quartet of transcriptional regulators, each one preferentially activating a specific subset of pluripotency factors. Among these, NANOG and MYC have the requirement of TGF-beta for self-renewal.

Therefore, we asked whether forced expression of the four factors would maintain pluripotency also in EpiSCs. We generated both GOF1827 and OEC2 38 EpiSCs stably expressing the four transcription factors (Supplementary Fig. 4a). TGF-beta inhibition led to a reduction of Nanog, Oct4, Otx2 and Fgft5 (Supplementary Fig. 4b) and none of the four factors were able to maintain the expression of the markers analysed, with the exception of Otx2 maintained only by KLF7. We conclude that the ability of NANOG, KLF7, MYC and ZNF398 to maintain pluripotency is not conserved in murine EpiSCs.
been extensively investigated as pluripotency regulators\(^{10,11,19,39}\). KLF7 is a Kruppel-like factor and other members of the same family, such as KLF2/4/5, are known regulators of pluripotency\(^{40}\). Conversely, ZNF398 has never been implicated in regulation of pluripotency, prompting us to choose it for further molecular characterisation.

ZNF398 represses differentiation and mesenchymal genes. When TGF-beta is blocked hPSCs lose pluripotency and undergo a morphological change. After focusing on the pluripotency regulators (Figs. 3 and 4), we decided to study the global effect of TGF-beta on hPSC function. Thus, we performed an unbiased transcriptional analysis and observed that upon SB43 treatment
538 genes were downregulated and 717 were upregulated (Fig. 5a). Gene Ontology (GO) enrichment analysis identified several categories associated with cell adhesion, epithelial to mesenchymal transition and organisation of the extracellular matrix, in agreement with the observed morphological change (Fig. 5b). Among them we identified a subset of genes specifically associated with epithelial character, that were downregulated by SB43 (Fig. 5c, epithelial). Moreover, we observed several gene categories associated with formation and function of neural cells (Fig. 5b), corresponding to a set of genes upregulated by SB43 (Fig. 5c, neuroectodermal). Upregulation of neuroectodermal genes was expected from studies performed in different model systems showing that TGF-beta, Activin A and Nodal block neuroectoderm formation41,42. Indeed, inhibition of TGF-beta is commonly used for neuroectodermal differentiation protocols43. Next, we asked whether the forced expression of ZNF398 was able to counteract such transcriptional changes and observed a reduction in neuroectodermal genes and boosted expression of epithelial genes (Fig. 5c). We conclude that ZNF398 is activated by TGF-beta to maintain the correct expression of neuroectodermal and epithelial genes in hPSCs.

Next, we asked whether ZNF398 would be required to control TGF-beta-dependent transcriptional programmes. We performed siRNA-mediated knockdown of ZNF398 and observed no effect on self-renewal (Supplementary Fig. 5a, b), as expected from the presence of four factors that are individually able to maintain pluripotency downstream of TGF-beta. However, ZNF398 knockdown during the early phases (5 days) of differentiation resulted in further reduction of pluripotency and epithelial markers, and enhanced induction of neuroectodermal genes (Fig. 5d, see also Supplementary Fig. 5c) to an extent comparable or greater than NANOG knockdown.

To further investigate the capacity of ZNF398 to regulate pluripotency and the epithelial character of hPSCs in an independent assay, we performed embryoid bodies (EBs) differentiation. Forced expression of ZNF398 was able to activate expression of pluripotency and epithelial markers, while repressing mesenchymal and germ layer markers relative to control cells (Fig. 5e). Collectively, these results indicate that ZNF398 promotes the expression of pluripotency and epithelial markers and represses genes associated with differentiation of hPSCs.

**ZNF398 activates transcription in concert with SMAD3.** We next sought to understand the molecular mechanism by which ZNF398 promotes pluripotency and epithelial character. ZNF398 contains several zinc-finger domains and it has been shown to recognise specific DNA sequences in COS-1 cells44. Therefore, we performed ChIP-seq for ZNF398 in two different hESCs lines and identified genomic regions bound by it containing a DNA motif similar to other ZNF factors (Supplementary Fig. 6a). Cooperative binding among transcription factors has been reported in several stem cell systems, thus we asked how similar the genome-wide-binding profile of ZNF398 is to those of other transcriptional regulators (data from CODEX45). Surprisingly, we found that ZNF398 clustered more closely with SMAD3 and the histone acetyl-transferase EP300, compared to the core pluripotency factors OCT4/NANOG or the Polycomb components (Fig. 6a, top panel). A similar analysis conducted on histone modifications-associated ZNF398 with regions decorated by acetylation of histone 3 on lysine 9 or 27 (Fig. 6a, bottom panel). Clustering results are confirmed by strong colocalisation of ZNF398, SMAD3, EP300 and H3K27ac (Fig. 6b). Histone acetylation is associated with both active promoters and enhancers, so we looked at the distribution of mono-methylation and tri-methylation of histone 3 on lysine 4, associated with active enhancers and promoters, respectively. We looked at ZNF398 peaks and found that 3595 ZNF398 peaks out of 5771 appeared as active enhancers (high levels of H3K4me1 and low H3K4me3), while the remaining 2176 peaks as active promoters (high H3K4me3). We conclude that ZNF398 preferentially colocalises with SMAD3 and EP300 at active enhancers and promoters in hPSCs.

The frequent colocalisation may be due to binding to neighbouring DNA regions or to physical interaction. A co-immunoprecipitation (Co-IP) assay indicates that SMAD3 and ZNF398 form a complex in hPSCs (Fig. 6c).

We observed that ZNF398 bound and activated the pluripotency factor LIN28B and the epithelial master regulator epithelial splicing regulatory protein 1 (ESRP1)46 (Fig. 7a), matching the pro-pluripotency and pro-epithelial activity of ZNF398. Interestingly, we also observed that LEFTY1, a known TGF-beta direct target, was also co-bound by ZNF398 (Fig. 7a). We therefore hypothesised that ZNF398 might potentiate the transcription of TGF-beta targets by binding SMAD3 targets. We functionally tested this hypothesis by comparing hPSCs expressing ZNF398 against control hPSCs. ZNF398 boosted the basal expression of LEFTY1 by >10 fold (i.e. in the presence of TGF-beta), and was even able to maintain residual LEFTY1 expression in the absence of TGF-beta signalling (Fig. 7b). We extended our analysis to all SMAD3 direct target genes (Fig. 1c) and observed that 23 out of 81 were also co-bound by ZNF398 (Fig. 7c, enrichment of 3.67 fold over those expected by chance, p-value = 3.49e-12, Chi-squared test). Importantly, the entire set of SMAD3-ZNF398 co-bound genes were significantly upregulated in cells expressing ZNF398 (Fig. 7d), further indicating a functional role of ZNF398 as activator of SMAD3 co-bound targets. This activity is specific...
for ZNF398, given that NANOG expression had no discernible effect on SMAD3-ZNF398 targets. Among the genes upregulated in hPSCs expressing ZNF398, we observed strong induction of several established direct targets of TGF-beta signal, such as LEFTY1/2, CER1, TGFBI1 and NODAL (Fig. 7e). We also analysed the dynamics of R-SMADs nuclear entry upon TGF-beta stimulation. Sixty minutes of treatment were sufficient to induce phosphorylation of SMAD3 and translocation from the cytoplasm to the nucleus (Supplemental Fig. 6b–d). Ectopic ZNF398 expression led to accelerated and enhanced nuclear translocation.

In sum, we conclude that ZNF398 colocalises with SMAD3 at active enhancers and promoters, activating the transcription of TGF-beta targets in hPSCs.

ZNF398 could be either a hPSC-specific or a general activator of the TGF-beta signal. We identified only two human cell lines expressing ZNF398 comparably to hPSCs (Supplementary Fig. 7a) and performed ZNF398 downregulation or over-expression, observing no differences in the induction of TGF-beta direct targets (Supplementary Fig. 7b, c). In two EpiSCs lines, stable expression of Zfp398—the mouse orthologue of ZNF398 mouse orthologue—we also observed no effect on the levels of TGF-beta targets (Supplementary Fig. 7d), in stark contrast with what we observed in hPSCs expressing ZNF398. We conclude that, among all the cell types we tested, ZNF398 activates the TGF-beta signal only in hPSCs.

ZNF398 is required for somatic cell reprogramming. So far, our results indicate that ZNF398 promotes the pluripotency and the epithelial character programmes in hPSCs. We decided to test the function of ZNF398 in an orthogonal system, the induction of pluripotency from somatic cells. Reprogramming from somatic cells, such as fibroblasts, requires an early mesenchymal to epithelial transition (MET) followed by the activation of endogenous pluripotency factors. We noticed that ZNF398 is expressed in human fibroblasts (Supplementary Fig. 8a), raising the possibility that ZNF398 promotes acquisition of epithelial character and pluripotency from early stages of reprogramming. We reprogrammed human fibroblasts by delivery of mRNAs encoding for either OSKMN1, OCT4, SOX2, KLF4, MYC, NANO5, LIN28A) or OSKM, in combination with siRNAs, allowing to test the requirement of endogenous ZNF398 for reprogramming (Fig. 8a). By day 6 of reprogramming, fibroblasts transfected with control siRNAs formed clusters of epithelial cells, such as keratinocytes, whereas ZNF398 knockdownexpression led to accelerated and enhanced nuclear translocation.

Among the genes upregulated for ZNF398, given that NANOG expression had no discernible effect on SMAD3-ZNF398 targets. Among the genes upregulated in hPSCs expressing ZNF398, we observed strong induction of several established direct targets of TGF-beta signal, such as LEFTY1/2, CER1, TGFBI1 and NODAL (Fig. 7e). We also analysed the dynamics of R-SMADs nuclear entry upon TGF-beta stimulation. Sixty minutes of treatment were sufficient to induce phosphorylation of SMAD3 and translocation from the cytoplasm to the nucleus (Supplemental Fig. 6b–d). Ectopic ZNF398 expression led to accelerated and enhanced nuclear translocation.

In sum, we conclude that ZNF398 colocalises with SMAD3 at active enhancers and promoters, activating the transcription of TGF-beta targets in hPSCs.
Fig. 7 ZNF398 boosts TGF-beta signal. a Left: Gene tracks of ZNF398 in two different hPSC lines (H9 and BGO1V), EP300, SMAD3, and H3K27ac. Middle: Barcharts of ChIP-qPCR for ZNF398 performed in BGO1V (dark grey) and H9 (light grey) cell lines on LIN28B, ESRP1 and LEFTY1 loci. Enrichment is expressed as a percentage of the DNA inputs. Bars indicate the mean of two biological replicates shown as dots. Right: RNA-seq of hPSCs (KiPS) stably expressing ZNF398 overexpressing cells, respectively. Absolute expression is reported as TPM. Source data are provided as a Source Data file. b LEFTY1 levels measured by qPCR of HE2 (light orange bars) and KiPS (dark orange bars) stably expressing an empty vector or ZNF398, untreated or treated with SB43 for 5 days. Bars indicate the mean ± SEM of independent experiments shown as dots (n = 4). Expression was normalised to the Empty SB43 samples and shown on a logarithmic scale. Unpaired two-tailed Mann-Whitney U test. Source data are provided as a Source Data file. c Pie-chart representing the 81 SMAD3 targets UP-regulated by TGF-beta induction identified in Fig. 1c. In dark blue are shown the 23 SMAD3 targets that are also bound by ZNF398. Such co-binding is significantly higher (p-value = 3.49e−12, Chi-squared test) than the one expected by chance, shown as the slice labelled with diagonal lines. d Mean-normalised expression levels of the 23 genes bound by ZNF398 and SMAD3. Shown data derived from RNA-seq analysis of KiPS stably expressing an empty vector, NANOG (serving as a control) or ZNF398. For each gene, data was normalised to the mean-expression across the three samples. Box plot indicates 25th, 50th and 75th percentile; whiskers indicate minimum and maximum. Unpaired two-tailed t-test. e Scatter plot showing RNA-seq data from KiPS stably expressing an empty vector (Empty) or ZNF398. DOWN-regulated (Log2FC < −1) and UP-regulated (Log2FC > 1) genes are indicated in blue and orange respectively.

Discussion
TGF-beta signalling is critical for hPSC self-renewal5–7. The transcription factor NANOG was first identified in murine ESCs for its capacity to maintain pluripotency in the absence of exogenous signals16. Such activity was also found conserved in hPSCs and it was shown that TGF-beta directly induces NANOG expression in hPSCs10,11. However, an unbiased and systematic analysis of
TGF-beta functional mediators in hPSCs was still missing. For this reason, we performed a transcriptome-level analysis of TGF-beta targets followed by a gain-of-function screening to identify uncharacterised pluripotency regulators.

Loss-of-function screenings have been performed in hPSCs, whereby genes were inactivated by RNA interference or using the CRISPR system. Such studies identified some critical pluripotency regulators, such as PRDM14 or BCOR. However, loss-of-function approaches might fail to identify critical regulators because of functional redundancy with other factors. For example, a CRISPR screening in murine ESCs failed to identify the majority of known pluripotency factors, likely because the pluripotency network is highly redundant and robust to inactivation of single factors. For this reason, we chose a gain-of-function screening approach, whereby individual putative pluripotency regulators are exogenously expressed in hPSCs and their capacity to maintain pluripotency is tested. Such an approach allowed the identification of several critical murine pluripotency regulators.
We identified a quartet of transcription factors, NANOG, MYC, KLF7 and ZNF398, which individually promote hPSC self-renewal. Interestingly, each of these four factors activates a specific subset of human pluripotency regulators,3,19,29–34, indicating that the human pluripotency network is flexible and can be maintained under different configurations. Among them, ZNF398 controls both pluripotency and epithelial genes downstream of TGF-beta (Fig. 8g).

Our analyses identified an extended set of functional human pluripotency regulators beyond the core factors OCT4, SOX2 and NANOG (Fig. 4b). It will be interesting to apply computational modelling20 to reconstruct the network of interactions among such factors in order to study how such a network reconfigures itself after perturbations or during reprogramming.

Interestingly, only a fraction of human pluripotency regulators are robustly expressed in murine ESCs (data from ref. 21). This observation is in part attributable to differences in the developmental stage, as conventional hPSCs are in a pluripotent stage primed for differentiation, whereas murine ESCs are in a more primitive, naïve state of pluripotency.20,37

However, naïve hPSCs have been recently obtained21–24 and we observed that ZNF398 and KLF7 are robustly expressed in hPSCs regardless of their pluripotent state. Moreover, forced expression of both genes could not maintain pluripotency in primed EpilSCs (Supplementary Fig. 4) and KLF7 expression in murine ESCs had no effect40, indicating that the two factors are human-specific pluripotency regulators.

It will be interesting to test whether the functions of TGF-beta and its direct targets are conserved or divergent in naïve and primed hPSCs.

Inhibitors of differentiation (ID) genes, such as ID1, block neural differentiation in the developing mouse embryo51 and in murine pluripotent stem cells52. ID1 is induced by BMP and by TGF-beta,52,53 also shown in our experiments. ID1 expression had a mild yet reproducible effect on AP-positive colony formation and maintenance of PRDM14 (Fig. 3a), and in the future it will be interesting to study whether ID1 inhibits neural differentiation also in hPSCs.

ZNF398 is a member of the Krüppel-associated box domain zinc finger proteins (KZFPs), the largest family of transcriptional regulators found in higher vertebrates. The majority of the 350 KZFPs identified in humans have been found to be associated with repression of transposable elements,54, playing key roles during early embryogenesis. Interestingly, roughly one-third of KZFPs, were found to be associated with gene promoters, as in the case of ZNF398.

We are tempted to speculate that some members of such a large family might have acquired new roles, beyond silencing of transposable elements and in so doing, contributed to the evolution of gene-regulatory networks.

ZNF398, also known as ZER6, has never been implicated in regulation of pluripotency. Previous studies reported that ZNF398 directly activates transcription44 and is regulated by Oestrogen Receptor Alpha44,55. Two isoforms of ZNF398 have been described,44,55,56 called p71 and p52. The shorter isoform (p52), lacks a N-terminal domain and promotes proliferation of cancer cells by ubiquitination of p5336. In hPSCs the longer isoform (p71) is predominant and has been used in all our experiments. It will be interesting to test whether p52, which lacks the N-terminal domain, regulates pluripotency in hPSCs.

Our results have also potential implications for reprogramming: ZNF398 knockdown strongly reduced reprogramming efficiency, indicating a critical role during establishment of pluripotency.

In particular, we observed reduced morphological conversion from mesenchymal to epithelial-like cells and reduced expression of epithelial markers and pluripotency markers, further indicating that ZNF398 promotes both pluripotency and epithelial character.

It will be interesting to see if ZNF398, or other members of the extended set of human pluripotency regulators, can be used to generate iPSCs at higher efficiency or to identify fully reprogrammed cells.

Methods

Cell culture. hESCs (HEK2, H9 and BG01V/hOOG [BG01V, Gibco R7799105]) and hiPSCs (KIPS, Keratinocytes induced Pluripotent Stem Cells) were cultured in feeder-free on pre-coated plates with 0.5% growth factor-reduced Matrigel (CORNING 356251) (vol/vol in PBS with MgCl2/CaCl2, Sigma-Aldrich D8662) in E8 medium (made in-house according to Chen et al.57 or in mTeSR (StemCell Technologies 05850) at 37 °C, 5% CO2, 5% O2. Cells were passaged every 3–4 days at a split ratio of 1:8 following dissociation with 0.5 mM EDTA (Invitrogen AM99260G) in PBS without MgCl2/CaCl2 (Sigma-Aldrich D8662), pH 7.4. The human foreskin fibroblasts B) (passage 12, ATCC, CRL-2522) were cultured in DMEM/F12 (Sigma-Aldrich D6421) with 10% foetal bovine serum (FBS; Sigma-Aldrich E7524) at 37 °C, 5% CO2, 21% O2. The H9 line (WA09) was obtained from and used under authorisation from WiCell Research Institute. The KIPS line was derived by reprogramming of human keratinocytes21 (Invitrogen) with Sendai viruses encoding for OSKMs and kindly provided by Austin Smith’s laboratory. The HES2 line was derived from a female human embryo at the blastocyst stage, as described in ref. 37 and kindly provided by Nicola Elvassore’s laboratory. Episomics (GOF1827 and OF2C28, kindly provided by Hans R. Schöler’s laboratory and Austin Smith’s laboratory, respectively) were cultured on serum-coated (GMEM [Sigma-Aldrich G5154] with 10% FBS) plates in serum-free media N2B27 (DMEM/F12 [Gibco 11320-074], and Neurobasal in 1:1 ratio [Gibco 21103-049], with 1:200 N2 Supplement [Gibco 17502-048], and 1:100 B27 Supplement [Gibco 17504-044], 2 mM L-glutamine [Gibco 25030-024], 0.1 mM 2-mercaptoethanol [Sigma-Aldrich M3148]) supplemented with FGF2 (12 ng/mL, QKINE 0002, recombinant zebrafish FGF2) and Activin A (20 ng/mL, QKINE 0001), and passed as small cell clumps every 2 days.
MCF10A and MCF10neo1 were cultured in DMEM/F12 with 5% horse serum (HS) (ThermoFisher 16050-122), 10 µg/ml insulin (Sigma-Aldrich B9278), 100 ng/ml cholera toxin (Sigma-Aldrich C6202), 20 ng/ml HGF (PeptidePro, AF100-151), 500 ng/ml hydrocortisone (Sigma-Aldrich H9369) and 2 mM l-glutamine. RPE-1, MCF10Ca1a, A549 and MDA-MB-231 were cultured in DMEM/F12 with 10% FBS and 2 mM l-glutamine. HEK293T and HaCaT were cultured in DMEM ( Gibco 41965-039) with 10% FBS and 2 mM l-glutamine. WI-38 cells were cultured in MEM (Gibco 32340-026) with 10% FBS and 5% O2. HepG2 were cultured in MEM with 10% FBS, 1.5% MEM non-essential amino acids (NEAA, Invitrogen 1140-036) and 4 mM l-glutamine. MCF10A, MCF10AnoeT, RPE-1, MCF10Ca1a, A549, MDA-MB-231, HEK293T, WI-38 and HepG2 were kindly provided by Sirio Dupont laboratory. HaCaT cells were kindly provided by Stefano Piccolo’s laboratory.

All cell lines were mycoplasma-negative (Mycoalert, Lonza).

### Treatment with inhibitors and cytokines

Treatments were performed either under feeder-free conditions or on feeders (MEF, Murine Embryonic Fibroblasts mitotically inactivated, DR4 ATCC). For the validation experiments of Fig. 2a in feeder-free, KiPs were plated on plastic coated with 0.5% Matrigel. The next day, cells were treated with DMSO (Sigma-Aldrich D2650) or 10 µM SB43 (Axon Medchem, 1509) overnight. The morning after, cells were treated with 2 ng/ml of TGFß1 (Peprotech 100-20) or with 25 ng/ml of Activin A. For the BMP induction experiment in Supplementary Fig. 6c, KiPs were plated under feeder-free conditions. The next day, cells were treated with DMSO on 0.1% NEAA and 0.1 mM 2-mercaptoethanol and with 10 ng/ml FGF2. The next day, cells were treated with DMSO or with 10 µM SB43 overnight. The morning after, cells were treated with 2 ng/ml of TGFß1 (Peprotech 10-20) or with 25 ng/ml of Activin A.

#### Generation of hPSCs stably expressing genes of interest

In order to generate the plasmid, the candidates (NANO2, ZNF398, KL7, MYC, ETS2, OTX2, ID1, BCL2, and 3′ end of pENTR2B donor vector). Then, the transgenes were Gateway cloned into the same destination vector containing PB-CAG-DEST-bGHpA and PGK-Hygro selection cassette.

#### Murine Episc experiments

For generation of stable transgenic lines over-expressing candidate genes, Episc were reverse-transfected with 3 µl of Lipo- fectamine 2000 (Invitrogen 11668-019) using 500 ng of PB transposon plasmid harbouring the indicated factor and 500 ng of transposase in 200 µl of Opti-MEM (Gibco 51985-026). 1.2 x 10^5 cells in 800 µl of N2B27 with FGF2 (12 ng/ml) and Activin A (20 ng/ml) and 10 µM ROCKi were added to the transfection mix and plated in serum-coated 2-well plate (2 cm²). The next day the medium was changed and Hygromycin B (200 µg/ml; Invitrogen 108780710) was added after 48 h. For the overexpression experiments, hPSCs stably expressing an empty vector or the candidates were plated. The next day, cells were treated with DMSO or 10 µM SB45 for 5 days and then analysed as indicated in Supplementary Fig. 2a.

#### siRNA and DNA transfection in HEK293T and HaCaT cells

Cells were plated at 20% confluency on a 24-well plate the day before transfection. For transfection with siRNAs, each individual well was transfected with Lipofectamine RNAiMAX reagent (ThermoFisher 13770-753) following the manufacturer protocol.

For transfection with DNA, each individual well was transfected with 5 ng of pCMV-Tag (Sigma-Aldrich T7-7641) in 1 ml of E8 medium with 10 µM ROCKi after 4 h. siRNAs were transfected at a final concentration of 20 nM using Stemfect™ RNA transfection Kit (STEM- GENT 00-0069), following the protocol for forward transfection.

For a 24-well plate (2 cm²), we used 0.52 µl of transfection reagent, 2 µl of 10 µM siRNA solution and 25 µl of transfection buffer. After waiting 20 min, we mixed the transfection mixture with 1 ml of E8 medium. The medium was changed after overnight incubation. See Supplementary Table 1 for sequences of the siRNAs used.

#### Ebs differentiation assay

KiPs stably expressing an empty vector or ZNF398 were detached as clumps with EDTA and plated on ultra low attachment surface (CORNING 3437) in E8 medium with 10 µM ROCKi. After 2 days, E8 medium was substituted with DMEM, 20% FBS, 1% NEAA, 1% NEAA and 0.1 mM 2-mercaptoethanol. Medium was changed every 2 days.

#### Reprogramming

All reprogramming experiments were performed in microfluidics in hypoxia conditions (37 °C, 5% CO2, 5% O2)45. The protocol for reprogramming experiments was optimised to transfect siRNA in order to test the requirement of ZNF398 for reprogramming.

Briefly, microfluidic channels were coated with 25 µg/ml Vitronectin (ThermoFisher, A14700) for 1 h at room temperature (RT). In the case of OSKMN reprogramming, fibroblasts were seeded at day 0 at 30 cells/mm² in DMEM/10% FBS. On day 1, 9 h before the first mRNAs transfection, we applied E6 well plate in a standard hPSC culture. On day 3, the media were replaced with DMEM without serum and with 10 µM SB43 or 5 µM ROCKi, 0.1 µM LSD1i (RN-1, EMD Millipore 498749) and 20% KSR (Gibco, 10820828). The transfection mix was prepared according to the StemMACS™ RNA Transfection Kit (Miltenyi Biotec, 130-104-463) and Stergent StemRNA-NM Reprogramming Kit (Repocal, 00-0076) (OSKMN not-modified RNA (NM) and EK8 NM and treated into reduced interferon and we prepared the RNA mix according to the manufacturer’s instructions.

In the case of OSKM reprogramming, individual modified mRNAs (OCT4, SOX2, KLF4 and MYC) were made in-house by in vitro transcription using mRNA synthesis with HiScript™ T7 ARCA mRNA Kit (NEB E2060S) according to the manufacturer’s instructions. OCT4, SOX2, KLF4 and MYC were amplified from cDNA and cloned into the pENTR2B donor vector.

#### Immunofluorescence and stainings

Immunofluorescence analysis was performed on 1% Matrigel-coated glass coverslip in wells or in situ in microfluidic channels with the same protocol. Cells were fixed in 4% formaldehyde (Sigma-Aldrich Aldrich 78775) in PBS for 10 min at RT, washed in PBS, permeabilized for 1 h in PBS with 0.3% Triton X-100 (PBST) at RT, and blocked in PBST with 5% of HS. For alkaline phosphatase staining, cells were fixed with a 4% formaldehyde solution and stained using an alkaline phosphatase detection kit (Sigma-Aldrich 86R-1KT). Plates were scanned using an Epson scanner and scored manually.

#### Image analysis

 Fiji 1.0 (ImageJ)28 was used for image analysis. Fluorescence intensity across hPSCs (Supplementary Fig. 6b) was measured using the Plot Profile function. For each condition, 48 cells from six randomly selected fields were analysed. Fluorescence intensity (Fig. 4a, Supplementary Fig. 3c) was quantified using Cell Profiler software (v3.1.8).

#### Western blotting

To monitor endogenous protein levels, cells were detached, medium removed and frozen at –80 °C prior to processing. Pellets were then thawed and resuspended in 10 ml/cell l l HPO buffer (50 nM Hepes pH 7.5, 100 mM NaCl, 50 mM KCl, 1% triton X-100, 0.5% NP-40, 5% glycerol, 2 mM MgCl2) freshly supplemented with 1 mM DTT, protease inhibitors (Roche 39802300) and phosphatase inhibitors (Sigma-Aldrich P5726). Western blotting was performed as in ref. 39.
Western blotting was acquired with LAS400 ImageQuant 1.2. Antibodies were detected in Supplementary Table 2. Unsupervised gels are provided in the Source data file.

**Quantitative PCR.** Total RNA was isolated using Total RNA Purification Kit (Norgen Biotech 37500), and complementary DNA (cDNA) was made from 500 ng using M-MLV reverse transcriptase (Invitrogen 28025-013) and dN6 primers. For real-time PCR SYBR Green Master mix (Bioline BIO-94020) was used. Primers are detailed in Supplementary Table 3. Three technical replicates were carried out for all quantitative PCR. GAPDH was used as endogenous control to normalise expression. qPCR data were acquired with QuantStudio® 6F8 Flex Software 1.0.

**RNA sequencing.** For induction experiments (Fig. 1d), poly(A) mRNA was purified from total RNA using Dynabeads mRNA direct kit (ThermoFisher, 61011). Quantity and quality of the starting mRNA were checked by Qubit and Agilent Bioanalyzer 2100 RNA pico chip. The template library was prepared according to the manufacturer recommendations. Sequencing on the Illumina NextSeq 500 platform was performed using the Illumina NextSeq 500 sample Prep Kit (Illumina) following the manufacturer’s instructions. Sequencing of RNA samples was performed on the Illumina NextSeq 500 platform. Reads were mapped to the Homo sapiens hg19 reference assembly using Bowtie (v1.1.2), and Bowtie2 (v2.0.6). Raw reads were quality checked and downsampled (75 bp reads) and then computationally mapped using edgeR package (v3.4.2)60 and R (version 3.5.2. R Core Team 2018). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. [https://www.R-project.org/](https://www.R-project.org/). EdgeR fits generalise negative binomial generalised linear models and conducts likelihood ratio test. Raw counts were normalised using edgeR package (v3.2.4)60 by library size and reads per kilobase per million mapped reads (RPKM).

The enrichment of the template library was achieved using the Ion OneTouch 2 enrichment system (ThermoFisher). Ion Proton sequencer and IP2 chip were prepared according to the manufacturer’s recommendations. Raw reads were aligned in two steps: first reads were aligned on genome build GRCh37.p13 with STAR (v2.4), reads that were not aligned in this step were realigned with bowtie2 (v2.1.0) aligned in two steps:

Preparation of RNA sequencing library sample was achieved using the Ion Total RNA-Seq Kit v2 (ThermoFisher, 4475936). Quantity and size distribution of the library were analysed using the Agilent Bioanalyzer 2100 DNA HS chip. Emulsion PCR using 10 ml of 100 ml PM library was performed using a OneTouch 2 instrument (ThermoFisher, 4474778) with an Ion PI Template OT2 200 kit following the manufacturer’s instructions (ThermoFisher, 4488318). The enrichment of the template library was achieved using the Ion OneTouch 2 enrichment system (ThermoFisher). Ion Proton sequencer and IP2 chip were prepared according to the manufacturer’s recommendations. Raw reads were aligned in two steps: first reads were aligned on genome build GRCh37.p13 with STAR (v2.4), reads that were not aligned in this step were realigned with bowtie2 (v2.2.4). Raw counts over the ensemble annotation release 75 were obtained with htsSamtools (v0.1.18). Read counts were normalized using the library size and the total number of reads (75 bp reads) and were extended to the average insert size (150 bp), and identical reads (reads starting and ending at the same positions) were collapsed. Peak calling was performed using MACS v2 (v1.4.2), selecting only peaks with q-value < 0.05. A non-redundant common set of peaks between the two ZNF398 ChiP-seq replicates was generated using MACS v2 (v2.0.10)66, selecting only peaks that were called in both replicates. Only peaks that were resided to ±200 bp surrounding their center and motif discovery was performed using MEME (v4.10.1)70. For correlation analyses and comparison of ZNF398 genome occupancy with known factors/histone modifications, data was collected from the GEO database for the following datasets: GSE54471 (H3K27ac and H3K4me3, Homo sapiens, GSE54471_1), GSE54601 (H3K79me3, Homo sapiens, GSE54601_1), GSE18325 (H3K9me3), GSE37725 (NANOG). Data for POUF1 and EP300 was instead obtained from the ENCODE database ([https://www.encodeproject.org/](https://www.encodeproject.org/)). All samples were analysed as stated above. Spearman correlations between genomic occupancy profiles were computed using the multibamSummarizer and heatmapCorrelation utilities from deepTools V2.2.4. Heatmaps of peak densities around ZNF398 peaks centers were generated using in-house developed scripts.

For SMAD3 and SMAD2 ChiP-qPCR, HESC samples (H9 and BG01V) were treated with 25 ng/ml Activin A to activate the TGF-β pathway for 1 h and crosslink the addition of formaldehyde for 1% for 10 min. RT-qPCR was carried out with 0.125 M MgCl₂ for 5 min at RT, and then washed with cold PBS. The cells were resuspended in isotonic buffer supplemented with 1% NP-40 to isolate nuclei. The pellets were resuspended in ChiP buffer (20 mM Tris–HCl pH 8.0, 0.5 mM EDTA, 1% SDS). Extracts were sonicated with the BioruptorH Twin (Diagenode) for 10 cycles (30 s on, 30 s off) and diluted with ChiP dilution buffer (20 mM Tris–HCl pH 8.0, 1% Triton X-100, 0.7% Na-Deoxycholate) and eluted in SDS elution buffer. De-crosslinked DNA was purified using QuickPCR Purification Kit (Qiagen) according to the manufacturer’s instructions.

ChIP-seq data were validated by ChIP-qPCR, using two independent biological replicates for each HESC sample (H9 and BG01V). The data represent qPCR measurements of the immunoprecipitated DNA performed using SYBR Green ER kit (Invitrogen) and were normalised to those obtained with a non-immune serum (IgG). The data are expressed as a percentage of the DNA inputs. Primers for ChIP–qPCR are detailed in Supplementary Table 4.

### Protein communoprecipitation.
To detect the protein interaction, nuclei were isolated from H9 cells expressing Avi-Tag-ZNF398 which were treated with 25 ng/ml Activin A for 1 h. Cells were lysed with isotonic buffer supplemented with 1% NP-40. The nuclei pellets were washed twice with IP buffer (20 mM Tris–HCl pH 8.0, 100 mM NaCl, 200 mM sucrose, 0.5 mM MgCl₂, 5 mM CaCl₂, 5 μM ZnCl₂) and were treated with micrococcal nuclease at 30°C for 10 min. Nuclear proteins were incubated with 2 μg of indicated antibodies (Supplementary Table 2) overnight at 4°C. The immunoprecipitated complexes were incubated with Protein G magnetic beads (Invitrogen) for 2 h at 4°C and then were washed three times with IP buffer plus 0.5% NP-40. The precipitated proteins were eluted by incubating with 0.5 M NaCl TE buffer and were further analysed by western blotting.

### Statistics and reproducibility.
For each dataset, sample size n refers to the number of independent experiments or biological replicates, shown as dots, as
stated in the figure legends. A Gaussian distribution was not assumed and \( p \)-values were calculated using the non-parametric unpaired two-tailed Mann-Whitney U test with correction for multiple comparisons. For the qPCR experiments (Fig. 2a, b) for which we used the unpaired two-tailed t-test, \( p \)-values were not calculated for datasets with \( n < 3 \). \( p \)-values are reported in the plots or figure legends. R software (v3.5.2) was used for statistical analysis.

All error bars indicate the standard error of the mean (SEM). All key experiments were repeated between two and five times independently, as indicated. Experiments of candidate’s functional validation were repeated using three different hPSC lines. All qPCR experiments were performed with three technical replicates.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** RNA-seq and ChIP-seq data for this study have been deposited in the Gene Expression Omnibus (GEO) database under the accession code: GSE133630. For the identification of TGF-beta transcriptional targets, we used available SMAD3 ChIP-seq data from Accession (Accession no. GSE21621), microarray data from Accession (Accession no. G-MEXP-1741) and RNA-seq data from H9 from Accession no. GSE24447, see Supplementary Fig. 1e. For correlation analyses and comparison of ZNF398 genome occupancy with known factors/histone modifications, data was collected from the GEO database for the following datasets: GSE5471 (H3K27ac and H3K4me1), GSE67084 (H3K27me3, H3K36me3, H3K9ac, SOX2), GSE118325 (H3K9me3), GSE473725 (NANOG). Data for POU5F1 and EP300 was instead obtained from the ENCODE database (https://www.encodeproject.org/). All plasmids, materials and data supporting the findings of this study are available from corresponding authors upon reasonable request. The source data underlying Figs. 1a, 2a, b, 3a, c, 4a, 5d, e, 6c, 7a, b, 8c, e and Supplementary Figs. 1b, f, g, 2d, 3a, c, d, 4a, b, 5c–6b, 7a–d, 8a–c are provided as a Source Data file.

Received: 12 September 2019; Accepted: 17 April 2020; Published online: 12 May 2020.

**References.**

1. Thomson, J. A. Embryonic stem cell line derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
2. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
3. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
4. Chen, G. et al. Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8**, 424–429 (2011).
5. Ludwig, T. E. et al. Derivation of human embryonic stem cells in defined conditions. *Nat. Biotechnol.* **24**, 185–187 (2006).
6. Vallier, L. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118**, 4495–4509 (2005).
7. Beattie, G. M. et al. Activin A maintains pluripotency of human embryonic stem cell in the absence of feeder layers. *Stem Cells** **23**, 489–495 (2005).
8. Chen, S. C., Shiao, A., Chin, A., & Oh, S. K. W. TGF-β2 and TGF-β1 are crucial for the maintenance of human embryonic stem cell proliferation on E6/E7 immortalized mouse embryonic fibroblasts. *J. Biotechnol.* **122**, 341–361 (2006).
9. Esieliovaa, L. et al. Comparative study of mouse and human feeder cells for human embryonic stem cells. *Int. J. Dev. Biol.* **52**, 353–363 (2008).
10. Vallier, L. et al. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* **136**, 1339–1349 (2009).
11. Xu, R.-H. et al. NANOG is a direct target of TGFβ-activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell** **3**, 196–206 (2008).
12. Ross, S. & Hill, C. S. How the Smads regulate transcription. *Nat. Rev. Mol. Cell Biol.* **13**, 1092–1101 (2011).
13. Nichols, J. & Smith, A. Naïve and primed pluripotent states. *Cell Stem Cell** **4**, 487–492 (2009).
14. Yang, J. et al. Stat3 activation is limiting for reprogramming to ground state pluripotency. *Cell Stem Cell** **7**, 319–328 (2010).
15. Chip, T. S. et al. MYC controls human pluripotent stem cell fate decisions through regulation of metabolic flux. *Stem Cell** **21**, 502–516.e9 (2017).
16. Yamane, M., Ohshuka, S., Matsuura, K., Nakamura, A. & Niwa, H. Overlapping functions of Krüppel-like factor family members: targeting multiple transcription factors to maintain the naïve pluripotency of mouse embryonic stem cells. *Development** **145**, dev16204 (2018).
17. Muñoz-Sanjuán, I. & Brivanlou, A. H. Neural induction, the default model and embryonic stem cells. *Nat. Rev. Neurosci.* **3**, 271–280 (2002).
18. Vallier, L., Reynolds, D. & Pedersen, R. A. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275**, 403–421 (2004).
19. Chambers, S. M. et al. Highly efficient neural conversion of human ES and iP cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* **27**, 275–280 (2009).
20. Conroy, A. T. et al. A novel zinc finger transcription factor with two isoforms that are differentially repressed by estrogen receptor-alpha. *J. Biol. Chem.* **277**, 9326–9334 (2002).
21. Sánchez-Castillo, M. et al. CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res.* **43**, D1117–D1123 (2015).
22. Warzecha, C. C. et al. An ESRP-regulated splicing programme is abrogated during the epiblast-mesenchymal transition. *EMBO J.* **29**, 3286–3300 (2010).
23. Warren, L. et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell** **7**, 618–630 (2010).
24. Gagliano, O. et al. Microfluidic reprogramming to pluripotency of human somatic cells. *Nat. Protoc.* **14**, 722–737 (2019).
25. Yilmaz, A., Peretz, M., Aharony, A., Sagi, I. & Benvenisty, N. Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat. Cell Biol.* **20**, 610–619 (2018).
50. Hackett, J. A. et al. Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. Nat. Commun. 9, 4292 (2018).

51. Lyden, D. et al. Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. Nature 401, 670–677 (1999).

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

53. Liang, Y.-Y., Brunicardi, F. C. & Lin, X. Smad3 mediates immediate early induction of Id1 by TGF-beta. Cell Res. 19, 140–148 (2009).

54. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 543, 550–554 (2017).

55. Stabach, P. R., Thiagarajan, M. M. & Weigel, R. J. Expression of ZER6 in Eμ-α positive breast cancer. J. Surg. Res. 126, 86–91 (2005).

56. Dang, V. C. et al. Zinc-finger protein p52 ZER6 accelerates colorectal cancer cell proliferation and tumour progression through promoting p53 ubiquitination. Ebiomedicine 48, 248–263 (2019).

57. Reubinoff, B. E., Pera, M. F., Tongson, A. & Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat. Biotechnol. 18, 399–404 (2000).

58. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).

59. Montagner, M. et al. Crosstalk with lung epithelial cells regulates Sfrp2-mediated latency in breast cancer dissemination. Nat. Cell Biol. 22, 289–296 (2020).

60. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. Oxf. Engl. 26, 139–140 (2010).

61. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).

62. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010).

63. Dennis, G. A. M. DAVIVI: database for annotation, visualization, and integrated discovery. Genome Biol. 4, R60 (2003).

64. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).

65. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).

66. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).

67. Kim, J., Cantor, A. B., Orkin, S. H. & Wang, J. Use of in vivo biotinylation to study protein–protein and protein–DNA interactions in mouse embryonic stem cells. Nat. Protoc. 4, 506–517 (2009).

68. Krepelova, A., Neri, F., Maldotti, M., Rapelli, S. & Oliviero, S. Myc and Max: stage-specific and tissue-specific transcription factors regulating cell proliferation and tumour progression through promoting p53 α positivity in breast cancer. Nat. Biotechnol. 19, W191 (2001).

69. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. Oxf. Engl. 26, 841–842 (2010).

70. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208 (2009).

71. Ramírez, F., Dündar, F., Diehl, S., Grünig, B. A. & Manke, T. DeepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191 (2014).

72. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283 (2011).