UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova
Department of Molecular Medicine

Ph.D. COURSE IN MOLECULAR MEDICINE
CURRICULUM: REGENERATIVE MEDICINE
SERIES XXXII

Dissecting the role of TGF-beta pathway in human Pluripotent Stem Cells

Thesis written with the financial contribution of Giovanni Armenise-Harvard Foundation

Coordinator: Prof. Stefano Piccolo
Supervisor: Prof. Graziano Martello

Ph.D. student: Irene Zorzan

INDEX

ABSTRACT

Human Embryonic Stem Cells (hESCs) and induced Pluripotent Stem cells (hiPSCs) are characterized by the ability to give rise all cell types found in the adult and to be expanded indefinitely *in vitro*. Understanding the molecular mechanisms controlling pluripotency is fundamental to differentiate human pluripotent cells into cells types useful for clinical applications. The signaling pathway of TGF-beta and FGF are known to maintain pluripotency in human cells. Only a handful of factors controlling pluripotency have previously been identified, such as the transcription factors OCT4, SOX2 and NANOG. Therefore, I used a systematic approach to identify novel components of the pluripotency network. Here I focused on the role of TGF-beta pathway, in order to find direct functional targets downstream of this pathway. Through comparative transcriptome analysis intersected with genome location data, I obtained a list of 21 putative transcription factors, out of which 8 were confirmed. Further functional assays led to the identification of four transcription factors that are able to maintain hESCs and hiPSCs undifferentiated in the absence of TGF-beta. Particularly, one of these four transcription factors has never been studied, so I focused on it. I then characterized the transcriptional program under the control of this factor in order to understand how it maintains the human pluripotency network. Interestingly, I found that this factor regulates both pluripotency and cell morphology. Finally, knockdown of this factor during the reprogramming strongly reduces the number of iPSCs obtained.

SOMMARIO

Le cellule staminali embrionali umane (hESCs) e le cellule staminali pluripotenti indotte (hiPSCs) sono caratterizzate dalla capacità di dare origine tutti i tipi cellulari presenti nell'adulto e di poterle espandere indefinitamente *in vitro*. Comprendere i meccanismi molecolari che controllano la pluripotenza è fondamentale per differenziare cellule pluripotenti umane in tutti i tipi cellulari utili per applicazioni cliniche. Le vie di segnalazione che mantengono la pluripotenza nelle cellule staminali pluripotenti umane sono TGF-beta e FGF. Ad oggi, sono stati identificati solo pochi fattori di trascrizione che controllano la pluripotenza, come i fattori di trascrizione OCT4, SOX2 e NANOG. Pertanto, ho utilizzato un approccio sistematico per identificare nuovi componenti del *network* di pluripotenza. Mi sono focalizzata sul ruolo di TGF-beta al fine di trovare target funzionali diretti che a valle di questa via di segnalazione siano in grado di mantenere lo stato di pluripotenza. Intersecando un'analisi comparativa del trascrittoma con dati relativi alla posizione nel genoma, ho ottenuto una lista di 21 fattori di trascrizione, di cui poi 8 sono stati confermati. Ulteriori test funzionali hanno portato all'identificazione di quattro fattori di trascrizione che sono in grado di mantenere hESCs e hiPSCs pluripotenti indifferenziate in assenza di TGF-beta. In particolare, uno di questi quattro fattori di trascrizione non è mai stato studiato, quindi mi sono focalizzata su di esso. Ho successivamente caratterizzato il programma trascrizionale controllato da questo fattore per capire come sia in grado di mantenere la pluripotenza. È interessante notare che questo nuovo fattore regola sia la pluripotenza che la morfologia cellulare, ossia l'identità epiteliale. Infine, il knockdown di questo fattore durante la riprogrammazione somatica riduce fortemente il numero di colonie di iPSCs ottenute.

INTRODUCTION

Pluripotency *in vivo* and *in vitro*

In mammals, the fertilized oocyte undergoes several rounds of cleavage divisions and the resulting cells are called blastomeres. Subsequently, the blastomeres compact each other generating a cell complex called morula. At 32-cell stage there is the formation of the blastocyst. The blastocyst consists of 3 cell types: the trophoblast that is fated to form the placenta and provides nutrients to the embryo, the hypoblast that will give rise to the visceral endoderm (also known as primitive endoderm) and the epiblast that holds the ability to generate the entire embryo (Boroviak and Nichols, 2017; Martello and Smith, 2014; Nichols and Smith, 2009, 2011). The epiblast is a cluster of 10-20 undifferentiated cells that are placed between the trophoblast and the hypoblast. Importantly, epiblast cells are characterized by the ability to give rise all differentiated cells of the three germ layers, namely ectoderm, mesoderm and endoderm. This ability is called pluripotency. Epiblast cells can be expanded indefinitely *in vitro* maintaining their pluripotency. This ability is defined as self-renewal and, along with the capacity to differentiate, are the defining features of stem cells. Thus, epiblast is a transitory state that, under precise culture conditions, can be captured *in vitro* as what we commonly call embryonic stem cells (ESCs - Martin, 1981; Figure 1A).

An elegant experiment to prove the pluripotency capacity of ESCs is to inject them into an early blastocyst. If ESCs are able to colonize the blastocyst and to differentiate *in vivo* generating chimeric animals, this means that they are pluripotent (Martello and Smith, 2014).

Human Pluripotent Stem Cells (hPSCs)

Human Pluripotent Stem Cells (hPSCs) have been derived from the epiblast of pre-implantation blastocyst as human Embryonic Stem Cells (hESCs, Thomson, 1998) or from somatic cells via transcription factors-mediated reprogramming as induced pluripotent stem cells (hiPSCs, Takahashi et al., 2007; Yu et al., 2007, Figure 1A). The

majority of hESCs were isolated from supernumerary and cryopreserved embryos generated during *in vitro* fertilization procedures. The huge benefit in using hiPSCs rather than hESCs cells is they are not derived from embryos while showing comparable differentiation and self-renewal capacity. This aspect allows to overcome several ethical issues and legal limitations.

Human PSCs were initially cultured on layers of inactivated fibroblasts ("feeder cells"), that produce different adhesion and signaling molecules, and in a medium containing fetal bovine serum (FBS) (Chen et al., 2014; Thomson, 1998). More recently, culture conditions of hPSCs were optimized using a feeder-free culture condition consisting of a solubilized basement membrane preparation (*Matrigel*) to coat the cell culture plates and a chemically-defined medium (Chen et al., 2011; Johansson and Wiles, 1995; Ludwig et al., 2006; Vallier, 2005). These novel culture conditions allowed to identify the key signals that regulate pluripotency. Indeed, the main pathways involved in the maintenance of pluripotency in hPSCs are TGF-beta/Activin and FGF signaling (Levenstein et al., 2006; Vallier, 2005; Vallier et al., 2004). However, the mechanism of action of TGF-beta and FGF remain poorly characterized. In this study, I focused on the role of TGF-beta signaling.

Role of TGF-beta signaling in hPSCs

TGF-beta1 (TGFB1), Nodal and Activin A are growth factors that belong to the TGF-beta family and they interact with serine/threonine kinase receptors on the cell membrane. These growth factors bind to type II receptor (ACVRIIA/ACVRIIB/TGFβRII) allowing the recruitment, phosphorylation and activation of type I receptor (ALK4/5/7). Subsequently, the dimerization of the two receptors leads to the activation by phosphorylation of Receptor-SMADs (R-SMADs, mainly SMAD3 in hPSCs). Once phosphorylated, the R-SMADs form a complex that interacts with the co-SMAD (SMAD4) and then they migrate into the nucleus in order to regulate gene expression (Figure 1B). R-SMADs bind directly the DNA together with the histone acetyltransferase EP300 leading to activation of target genes. Moreover, R-SMADs are known to interact with several transcription factors and co-regulators that may

vary between different cell types, resulting in induction of transcriptional programs that are specific for the cell type (Dennler et al., 1998; Massagué et al., 2005; Ross and Hill, 2008; Shi et al., 1998; Zawel et al., 1998).
In hPSCs, TGF-beta signaling has been shown to actively promote self-renewal by stabilizing the expression of NANOG, SOX2 and OCT4 (Vallier et al., 2009; Xu et al., 2008), three transcription factors that are critical to maintain pluripotency both in the embryo or in pluripotent cells cultured *in vitro* (Masui et al., 2007; Silva et al., 2009). Furthermore, TGF-beta signaling blocks differentiation towards neuroectoderm (Vallier et al., 2004). For this reason, inhibition of TGF-beta is conventionally used for neuroectodermal differentiation protocols (Chambers et al., 2009). Nevertheless, the downstream mediators of TGF-beta and the molecular mechanism by which they sustain the pluripotency network and block differentiation are still unclear (Figure 1C).

Main pluripotency factors

A pluripotency factor can be defined as a gene product that specifically and directly sustains the maintenance of the ESCs identity. In other words, pluripotency factors are specific transcription factors that regulate and define the pluripotency state in ESCs. Essentially, there are three transcription factors that belong to what is called "core" of pluripotency: OCT4, SOX2 and NANOG. These transcription factors were initially identified in mouse ESCs (mESCs) and then confirmed also in hPSCs.
Oct4 (*POU-domain transcription factor Oct4*, also known as *Pou5f1*) was the first transcription factor defined as a regulator of pluripotency (Scholer et al., 1990). *In vivo*, Oct4 is expressed in both the oocytes and in the early embryo, while *in vitro* its expression was found in ESCs, in the embryonal carcinoma (EC) cells and in embryonal germ (EG) cells. Moreover, it has been shown that the deletion of Oct4 in the mouse embryo induces the differentiation of epiblast cells into trophectoderm (Nichols et al., 1998). This result was also confirmed in hESCs in which the deletion of *OCT4* results in differentiation towards trophectoderm (Babaie et al., 2007). Surprisingly, the overexpression of Oct4 in mESCs leads to differentiation (Niwa et al.,

2000). These data suggest that the regulation of the expression of Oct4 is crucial for the maintenance of pluripotency and self-renewal.

Another important pluripotency factor is SOX2 (*SRY-box transcription factor*). The inactivation of *Sox2* in mESCs results in trophoblast formation, similarly to *Oct4* deletion (Masui et al., 2007). Sox2 interacts with Oct4, it binds the DNA with Oct4 at Oct/Sox elements and positively regulates the Oct4 transcription (Masui et al., 2007). Furthermore, the *SOX2* inactivation in hESCs induces differentiation towards trophectoderm and primitive endoderm (Adachi et al., 2010; Fong et al., 2008). In addition, the *SOX2* overexpression results in differentiation towards neuroectoderm suggesting that SOX2 could have a similar role of OCT4 (Wang et al., 2012).

Moreover, Oct4/Sox2 regulate the expression of an additional pluripotency factor that is *Nanog*. The loss of *Nanog* destabilizes pluripotent stem cells both *in vivo* and *in vitro* (Chambers et al., 2007; Silva et al., 2009). In hESCs, NANOG has been shown to maintain pluripotency in absence of a medium supporting pluripotent state (Darr et al., 2006) and to repress differentiation towards neuroectoderm (Vallier et al., 2009; Wang et al., 2012) and it has been demonstrated that the knockdown of *NANOG* results in differentiation toward extraembryonic endoderm and trophectoderm (Hyslop et al., 2005; Zaehres et al., 2005). More importantly, it has been shown that NANOG is induced by TGF-beta signalling (Vallier et al., 2009; Xu et al., 2008).

In the case of murine PSCs, additional factors have been identified, beyond Oct4, Sox2 and Nanog. Indeed, several studies characterised extensively the pluripotency network in the mouse system (Dunn et al., 2014; Martello et al., 2012, 2013). However, the pluripotency factors in hPSCs remain poorly identified. Recently, PRDM14 (Chia et al., 2010), LIN28B (Zhang et al., 2016), LIN28A (Yu et al., 2007) and MYC (Cliff et al., 2017) have been shown to play an important role in the maintenance and induction of pluripotency both in mouse and human PSCs. In particular, PRDM14 regulates directly the expression of OCT4 through its proximal enhancer (Chia et al., 2010). Moreover, the RNA-binding proteins LIN28B and LIN28A have been shown to increase the reprogramming efficiency (Zhang et al., 2016; Yu et al., 2007; see also

the paragraph “induction of pluripotency with somatic cell reprogramming”) and MYC has a critical role not only in the induction of pluripotency (Chappell and Dalton, 2013; see also the paragraph “induction of pluripotency with somatic cell reprogramming”), but also in the exit of pluripotency controlling metabolism (Cliff et al., 2017). However, all such factors are evolutionary conserved in mouse and human PSCs and we do not know whether human-specific pluripotency factors exist. Moreover, additional studies are needed to elucidate how human factors interact with each other in order to form a network controlling human pluripotency.

How to assess pluripotency in hPSCs with functional assays

In order to determine the pluripotency of a stem cell, several assays can be used. Regarding mESCs, the main assay is the chimera formation. As mentioned above, if a cell is pluripotent, when injected in a blastocyst, it will participate to the development of the host embryo (Nagy et al., 1990). For obviously ethical reasons, the chimera formation assay can not be employed in the human system. Thus, alternative assays were developed, such as the teratoma assay. A teratoma is a nonmalignant tumor composed of a mixture of cells of the three germ layers. In this assay, putative PSCs are implanted into an immunocompromised mice and if PSCs are pluripotent, they will produce the tumor containing derivatives of three germ layers (Bouma et al., 2017; Daley et al., 2009). However, teratoma assay need mices, high costs, it is time consuming and finally it is not a quantitative assay. So, other methods were employed to assess pluripotency *in vitro*. Indeed, the *in vitro* differentiation tool is the easier method to evaluate the pluripotency capacity of a stem cell. Typically, PSCs are induced to differentiate replacing medium that sustain pluripotency with a medium in which specific cytokines are added allowing a particular lineage differentiation. Alternately, when PSCs are cultured in a low attachment plate with a medium that neither maintains pluripotency nor induces a specific lineage differentiation, they grow in suspension generating three-dimensional (3D) aggregates, namely Embryoid Bodies (EBs, Figure 1D, Höpfl et al., 2004; Itskovitz-Eldor et al., 2000). Progressively, EBs cavitate and spontaneously differentiate into all

the 3 lineages (ectoderm, mesoderm and endoderm). However, EBs are disorganized structures that do not recapitulate the early phases of embryo development. This system is widely used not only to assess pluripotency but it is also applied as an initial step in several differentiation protocols. As a matter of fact, the addition in the medium of specific cytokines after a first step of EBs differentiation allow to differentiate PSC into specific mature differentiated cell types (Schuldiner et al., 2000).

Induction of pluripotency with somatic cell reprogramming

The identification of pluripotency factors allowed a great breakthrough: the induction of pluripotency in somatic cells, namely somatic cell reprogramming. Indeed, the forced expression of a specific cocktail of factors leads the conversion of somatic cells into the so-called Induced Pluripotent Stem Cells (iPSCs). In the first experiment performed by Yamanaka (Takahashi and Yamanaka, 2006; Takahashi et al., 2007), the factors present in the reprogramming cocktail were Oct4, Sox2, Klf4 and cMyc (OSKM). Klf4 is a known pluripotency factor characterized in murine ES cells (McConnell and Yang, 2010), whereas Myc has been added to increase transcription and proliferation during reprogramming (Chappell and Dalton, 2013). Induced PSCs were obtained from both murine and human somatic cells (Takahashi and Yamanaka, 2006; Takahashi et al., 2007), but in the case of human somatic reprogramming the protocol was optimized (Yu et al., 2007). In particular, Thomson and colleagues (Yu et al., 2007) independently identified a different cocktail of factors in which they replaced cMYC and KLF4 with NANOG and LIN28A. Subsequently, it has been shown that the combination of OCT4, SOX2, KLF4, cMYC, NANOG and LIN28A (OSKMNL) boosted the efficiency of iPSCs generation (Tanabe et al., 2013; Yu et al., 2009). Importantly, in the first protocols to generate iPSCs, the forced expression of pluripotency factors was driven by retroviral vectors that stably integrated into the genome potentially causing dangerous mutations. In order to solve this issue, alternative reprogramming methods have been developed such as Sendai virus (Fusaki et al., 2009; Nishimura et al., 2011), episomal vectors (Okita et al., 2011; Yu

et al., 2009), piggyBac vectors (Woltjen et al., 2009; Yusa et al., 2009), the delivery of recombinant proteins (Kim et al., 2009a; Zhou et al., 2009) or of modified messenger RNAs (mmRNAs, Luni et al., 2016; Warren et al., 2010). Specifically, the use of mmRNAs is currently the most efficient technology for integration-free reprogramming (Schlaeger et al., 2015). Furthermore, it has been recently shown that the efficiency of iPSCs generation can also increase by performing reprogramming in microfluidic system in which cells are in a confined microenvironment (Giulitti et al., 2018; Luni et al., 2016). Of note, the microfluidic system not only increases the reprogramming efficiency, but it allows also to drastically reduce the number of somatic cells and reagents (medium, mmRNAs and transfection reagent) needed.

Self-organization capacity of PSCs *in vivo* and *in vitro*

After implantation, epiblast cells self-organize into a polarized rosette-like structure to form the amniotic cavity (Deglincerti et al., 2016; Shahbazi et al., 2016), resembling the formation of pro-amniotic cavity in the mouse embryo (Bedzhov and Zernicka-Goetz, 2014). In particular, in human embryo, the epiblast cells in contact with trophoblast cells give rise the amniotic epithelium, while the epiblast cells in contact with the hypoblast form the epiblast disc (Rossant and Tam, 2017). After the formation of amniotic cavity, the epiblast cells ingress at the primitive streak to form the three germ layers (Rossant and Tam, 2017), but how this process occurs in the human embryo is still a black box.

Recently, it has been shown that hPSCs are characterized by the capacity to self-organize *in vitro* into three-dimensional (3D) structures, resembling the epiblast rosette-like structure of the early embryo (Figure 1E, Shahbazi and Zernicka-Goetz, 2018; Shahbazi et al., 2016; Shao et al., 2016, 2017; Taniguchi et al., 2015). Specifically, when a suspension of single cells are put in a Matrigel layer and they are cultured in a medium supporting pluripotency and containing components of the ECM (extracellular matrix), they form an epiblast-like structure composed by a polarized columnar epithelium and a lumen (Shahbazi and Zernicka-Goetz, 2018;

Shahbazi et al., 2019). However, the molecular mechanisms, the external signals and the mechanical cues that drive the process of self-organization are still unclear.
Interestingly, the epiblast-like structures have been also shown to be an ideal *in vitro* model to recapitulate the post-implantation human amniotic sac development (Shao et al., 2016, 2017) and the process of symmetry breaking (Simunovic et al., 2019) *in vitro*. More recently, the epiblast-like structures have been demonstrated to be able to mimic the specification of primitive streak cells and primordial germ cells (Zheng et al., 2019). Of note, such systems give the possibility to study *in vitro* early phases of embryonic development overcoming the ethical and technical limitations associated with the use of human embryos.

Mesenchymal-epithelial transition in development and reprogramming

Epithelial to mesenchymal transition (EMT) and its reverse process mesenchymal to epithelial transition (MET) are two fundamental and conserved mechanisms that direct morphogenesis and organogenesis. EMT implies the loss of epithelial polarity, cytoskeleton reorganization and cellular migration allowing the generation of the three germ layers and later the tissues and organs. On the contrary, MET is critical for the formation of epithelia at several stages of development. During MET, mesenchymal cells establish apical-basal polarity by a specific group of proteins and lipids, including polarity complex proteins, Rho GTPases and phosphoinositides leading the formation of tight junctions and the organization of cytoskeletal structures and organelles (Pei et al., 2019; Rodriguez-Boulan and Macara, 2014). In particular, epiblast cells both in mouse and in human express ECAD (*Epithelial-cadherin*, also known as *CDH1*, Cadherin 1) *in vivo* and, as mentioned above, when the amniotic cavity is formed, they start to acquire the features of polarized cells (Bedzhov and Zernicka-Goetz, 2014; Shahbazi et al., 2016). At the peri-implantation stage, epiblast cells express not only *ECAD*, but also other epithelial markers such as BETA-CATENIN (also known as *CTNNB1*), EZRIN (*EZR*), OCCLUDIN (*OCLN*) and the proteins that are involved in epithelial polarity such as PKC (*protein kinase C*), PARD3 and PARD6A. Interestingly, *in vitro* hPSCs form flat two-dimensional (2D) colonies

expressing epithelial markers such as ECAD and EPCAM (Epithelial Cell Adhesion Molecule) (Eastham et al., 2007; Lu et al., 2010). Conversely, it has been shown that when PSCs differentiate, they lose pluripotency and the epithelial identity and they start to express mesenchymal markers such as VIM (*Vimentin*), NCAD (*Neural-cadherin*), SLUG and SNAIL (Eastham et al., 2007). Moreover, the process of MET is essential during somatic reprogramming (Høffding and Hyttel, 2015; Shu and Pei, 2014). Indeed, fibroblasts are mesenchymal cells but during reprogramming, they change completely their morphology to become epithelial cells. Thus, the acquisition of an epithelial-like status is necessary to induce pluripotency. In the mouse system, it has been demonstrated that the transcription factor Klf4 induces the epithelial programme (Li et al., 2010). However, it is still unclear what factors regulate MET during reprogramming of human somatic cells.

AIM OF THE PROJECT

Our knowledge of the factors controlling pluripotency in human cells is by far more limited compared to what we know about the murine system. For instance, the molecular mechanism by which TGF-beta regulates the pluripotency network and the epithelial identity is still unknown. The aim of this project is to identify in a systematic way direct TGF-beta targets that act downstream of this pathway to sustain pluripotency and the epithelial character. By doing so I should be able to identify novel critical regulators of pluripotency, using an approach previously used to study the Wnt and LIF pathways in murine ES cells (Martello et al., 2012, 2013). Moreover, core factors are often included in reprogramming cocktails (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007), so the identification of novel human-specific pluripotency regulators could play a critical role for induction of human pluripotency.

First of all, I performed transcriptional analyses to identify novel TGF-beta targets in hPSCs and established a set of robust experiments (gain-of-function approach) to test the functional relevance of novel pluripotency regulators. In so doing, I identified 4 transcriptional regulators that maintain the human pluripotency network downstream TGF-beta. Among them, ZNF398 was never characterized before so I focused on it. I characterized the transcriptional program regulated by ZNF398 and I observed that ZNF398 promotes the expression of pluripotency and epithelial markers and represses genes associated with differentiation of hPSCs. Moreover, I found that ZNF398 acts as an activator with SMAD3 and EP300 and boosts the TGF-beta signaling. Furthermore, I tested the function of ZNF398 in the context of somatic cell reprogramming and found a drastic reduction in iPSC colonies upon ZNF398 knockdown. Finally, I further tested the functional role of ZNF398 in self-organization and I found that ZNF398 is epistatic to TGF-beta signal for the formation of 3D structures.

RESULTS

TGF-beta is required for self-renewal and epithelial identity in hPSCs

TGF-beta signaling is required for the maintenance of pluripotency and self-renewal of hPSCs (Ludwig et al., 2006; Vallier, 2005). Specifically, treatment with TGF-beta receptor inhibitor SB431542 (SB43) leads to differentiation.

I first wanted to independently validate such findings under the condition used in our laboratory. To do so, I expanded hPSCs under chemically defined conditions (Chen et al., 2011; Ludwig et al., 2006) and treated them with DMSO, as a control, or with SB43 for 5 days. Then, I performed Immunofluorescence protocol and analysed markers of pluripotency. Upon TGF-beta inhibition, the expression of the pluripotency factors POU5F1/OCT4 and NANOG (see Introduction) was downregulated (Figure 2A, right panel). To further confirm such results, I have also performed quantitative Polymerase Chain Reaction (qPCR) for the pluripotency markers (POU5F1/OCT4, NANOG and PRDM14) and observed strong reduction also at the level of mRNA (Figure 2C, top panel).

Human pluripotent colonies are composed of a monolayer of tightly packed cells, expressing epithelial markers (Eastham et al., 2007). SB43 treatment induces not only a morphological change with loss of cell-cell contact (Figure 2A, left panel), but also a reduction of the epithelial marker ECAD which is normally localised at the cell-cell boundary. Concomitantly, the mesenchymal marker VIM, a cytoskeletal component, was strongly upregulated (Figure 2B and Figure 2C, bottom panel), as previously described (Eastham et al., 2007).

These results confirm the essential role of TGF-beta in the maintenance of human pluripotency and epithelial identity.

Identification of TGF-beta transcriptional targets in hPSCs

I decided to study how TGF-beta controls gene programs associated with pluripotency and the epithelial character with an unbiased functional approach based on the identification of direct transcriptional targets followed by functional validation (Martello et al., 2012, 2013).

I reasoned that if a gene is a TGF-beta transcriptional target it should be bound by SMAD3 and either downregulated upon signal inhibition or rapidly induced upon stimulation. Thus, I first interrogated available Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq) data for SMAD3, in order to obtain a list of genes promoters bound by SMAD3 in human PSCs (Figure 3A, orange).

I then analysed transcriptome profiles for hPSCs in which TGF-beta has been inactivated for 48 hours and looked for genes that were downregulated (Figure 3A, light blue and Figure 3B, top panel yellow dots on the left). Intersecting the two gene lists, I identified 195 genes as candidates TGF-beta targets. In parallel, by intersecting SMAD3-bound genes with genes induced after 4 hours of acute stimulation, I identified 81 putative targets (Figures 3A, green and 3B, bottom panel yellow dots on the right). Several known TGF-beta direct targets, such as LEFTY1/2, SKIL and SMAD7, were identified (Figure 3B), supporting the validity of the strategy used.

I 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 programs. Finally, I included only genes robustly expressed (>3 RPKM in hPSCs) (Figure 3C), obtaining a list of 21 candidates. The threshold of 3 RPKM was chosen based on the fact that genes below such threshold were not detectable by other techniques such as qPCR.

To further validate the putative TGF-beta targets with an independent technique, I performed TGF-beta re-induction experiments and analysed them by qPCR. To do so, I treated hPSCs with SB43 overnight and the morning after I re-induced the TGF-beta pathway. In particular, I tested the responsiveness to two TGF-beta superfamily ligands - TGFB1 and Activin A - commonly used for hPSCs expansion and both acting

through SMAD3 (Chen et al., 2011; Johansson and Wiles, 1995; Ludwig et al., 2006; Vallier, 2005). I also tested whether targets were responsive to TGF-beta signal when cells were expanded either on feeders or under feeder-free conditions, given that TGF-beta signal maintains pluripotency under both conditions (Chen et al., 2006; Vallier et al., 2009; Xu et al., 2008).

To test if re-induction experiments worked correctly, I used some known TGF-beta targets, such as LEFTY1, SKIL and SMAD7 as positive controls (Shi and Massagué, 2003; Vallier et al., 2009; Xu et al., 2008). As expected, *LEFTY1*, *SKIL* and *SMAD7* were downregulated when TGF-beta was inhibited and they were upregulated upon re-induction (Figure 3D, grey box).

Moreover, some genes could be characterized by a different kinetic of expression: they could be induced more slowly or quickly than others and then their expression would tardily or rapidly decrease. Specifically, I noticed that *NANOG*, a known TGF-beta target playing a key role in the human pluripotency network (Vallier et al., 2009; Xu et al., 2008), was robustly induced only after 4 hours of treatment. I therefore conducted the re-induction experiments with an induction of TGF-beta for both 1 hour or 4 hours.

After extensive validation, I identified 8 genes (*ID1, MYC, BCOR, KLF7, OTX2, ZNF398, NANOG and ETS2*) as *bona fide* TGF-beta transcriptional targets in hPSCs, because they were consistently induced under multiple different conditions (Figures 3D and 3E). In figure 3E I presented three examples of direct TGF-beta targets: the known TGF-beta targets, SMAD7 and NANOG, and ZNF398, one of the novel TGF-beta targets identified in this work. As a matter of fact, these direct targets of TGF-beta are bound by SMAD3 (Figure 3E, see the peaks in the binding tracks of SMAD3) and, upon stimulation of TGF-beta, their expression is induced (Figure 3E, see qPCR for *SMAD7*, *NANOG* and *ZNF398* represented as bar charts).

Functional identification of pluripotency regulators in hPSCs

If a gene is a critical downstream mediator of TGF-beta signal in hPSCs, its forced expression should maintain pluripotency also when TGF-beta signaling is inhibited. In order to test the functional relevance of putative novel pluripotency regulators, I used a set of complementary experiments (Figure 4A).

I stably expressed putative regulators using piggyBac (PB) vectors that allow stable integration in the genome of mammalian cells. The PB vectors I used contains a strong CAG promoter, driving the expression of genes of interest. Moreover, the presence of a Hygromycin resistance gene in the PB vector, whose expression is driven by a PGK promoter, gave me the possibility to select only cells that had stably integrated the construct in the genome (see Methods for details). I performed both in hiPSCs (KiPS:Keratinocytes induced Pluripotent Stem cells) and in hESCs (HES2) overexpression of the putative regulators (ID1, MYC, BCOR, KLF7,OTX2, ZNF398, NANOG and ETS2). The reason for using both hiPSCs and hESCs is the known variability among different hPSC lines (Osafune et al., 2008). If a gene is a critical regulator of pluripotency, it should work consistently in different hPSC lines.
After Hygromycin selection, I plated overexpressing cells (and cells transfected with an empty vector as a control) with or without SB43 for 5 days and I analysed them with different assays (Figure 4A).

First, I performed a clonal assay, which allows to quantify the fraction of hPSCs able to self-renew and give rise to pluripotent colonies. I plated 2,000 cells in a well of a 12 well plate with or without SB43 and after 5 days I performed an Alkaline Phosphatase (AP) staining. Given the specific expression of this phosphatase in undifferentiated pluripotent cells, this assay is suitable to discriminate whether a cell is pluripotent or not. Only pluripotent cells are able to self-renew as single cells, giving rise to a colony that will be AP positive. In this way, I can quantify the fraction of pluripotent cells in a population. Cells transfected with an empty vector formed a reduced number of AP positive colonies when treated with SB43 (Figure 4B). Only

expression of NANOG, KLF7, MYC and ZNF398 fully rescued the formation of AP positive colonies in the presence of SB43, while other factors had only a partial, or no effect.

Second, I expanded cells at standard density and treated them with SB43 (Figure 4A). As shown in Figure 2A, SB43 treatment induces clear morphological changes associated with differentiation. In the case of forced expression of NANOG, KLF7, MYC and ZNF398 cells maintained a flat epithelial-like morphology, generally associated with pluripotency (Figure 4C), while other factors failed to do so and induced differentiation comparable to the empty vector.

Third, I evaluated the expression of *NANOG*, *OCT4* and *PRDM14* by qPCR and I confirmed that NANOG, KLF7, MYC and ZNF398 were able to maintain expression of pluripotency markers (Figure 4D) in presence of SB43, while all other candidates were not distinguishable from empty vector control. Importantly, similar results were obtained both with hESCs (Figure 4D, light green) and hiPSCs (Figure 4D, dark green). Moreover, I noticed that each candidate displayed specificity for different targets. For instance, ZNF398 and NANOG activated robustly *PRDM14* expression, while KLF7 fully maintained *OCT4* expression.
All these results indicate that forced expression of NANOG, KLF7, MYC and ZNF398 is individually sufficient to maintain pluripotency upon TGF-beta inhibition both in hESCs and iPSCs.

Additionally, I wanted to confirm these findings in another hESC line, the H9 cell line, because is the most used in the field of hPSCs (Löser et al., 2010). So, I overexpressed NANOG, KLF7, MYC and ZNF398 in H9 cell line and I performed the same experiments done with KiPS and HES2 cell lines. Also in H9 cell line, the forced expression of NANOG, KLF7, MYC and ZNF398 allowed formation of AP+ colonies in the absence of TGF-beta (Figure 5A). Furthermore, at the morphological level, the overexpression of NANOG, KLF7, MYC and ZNF398 was able to maintain undifferentiated colonies, even

though the effect of MYC and ZNF398 seemed less pronounced in H9 cell line compared to the results obtained in KiPS and HES2 cell lines (Figure 5B).

In gain-of function experiments, the overexpression effect of a putative target could be influenced by the level of transgene expression. In other words, the mild effect of MYC and ZNF398 observed in H9 could be due to low levels of expression. In order to test this possibility, I checked the level of transgenes expression in different cell lines. First, I checked the expression relative to endogenous levels of each candidate in the 3 lines, and observed that all factors were expressed between 4 and 7 times over the endogenous levels, with the exception of MYC and ZNF398 in H9, that were significantly lower (Figure 5C, left panel). Such results were confirmed using primers specific for the 3'UTR region of the transgenes, allowing detection of their exogenous expression (Figure 5C, right panel). Indeed, the procedure of transfection and selection was more troublesome in H9 cells. I conclude that the milder effect of MYC and ZNF398 in H9 could be due to lower expression in H9.

To further confirm the role of NANOG, KLF7, MYC and ZNF398 in the maintenance of pluripotency with an independent technique, I examined the expression of OCT4 and NANOG at the protein levels by Quantitative Immunostaining. Upon TGF-beta inhibition, the expression level of NANOG is completely abolished and OCT4 is reduced in the control cells (empty vector). On the contrary, the expression of NANOG and OCT4 is sustained in ZNF398 and KLF7 overexpressing cells despite the presence of TGF-beta inhibitor, and to a lesser extend in MYC overexpressing cells (Figure 6A for experiment performed in KiPS and 6B for experiment performed in H9). Of note, the NANOG expression is heterogeneous among individual cells in the overexpressing lines (KLF7, MYC and ZNF398 overexpressing cells), while OCT4 expression is much more homogeneous. This effect could be due to the variable level of transgene expression, because the overexpressing cell lines are not a clonal population.

Taken together, these results show the ability of NANOG, KLF7, MYC and ZNF398 overexpressing cells to maintain the undifferentiated state upon TGF-beta inhibition, although the effects of MYC are not consistent in different assays. Specifically, I noticed that MYC had a strong effect on AP+ colony formation and morphology despite the partial effects on OCT4 and NANOG expression (Figure 6A and 6B), suggesting that MYC might maintain pluripotency via other pluripotency factors. So, such discrepancy could be due to our limited knowledge of the factors regulating human pluripotency. I therefore decided to characterize the effects of expression of our candidates at the whole transcriptome level.

I performed transcriptome profiling by RNA-sequencing of the empty vector and NANOG, KLF7, MYC and ZNF398 overexpressing cells cultured with or without TGF-beta for 5 days. I searched for all genes highly expressed in hPSCs that were significantly reduced upon SB43 treatment for 5 days. In this way, I identified 538 genes downregulated that are associated with human pluripotency, including also *PRDM14* and *NANOG* (Figure 6C). Box plots show the expression of such genes in logarithmic scale. The mean reduction in expression is equal to 3.75 fold. All the 4 factors (NANOG, KLF7, MYC and ZNF398) were able to counteract the effects caused by TGF-beta inhibition (Figure 6C).

These results indicate that TGF-beta maintains pluripotency mainly via a quartet of transcriptional regulators. Among these, NANOG and MYC (Chappell and Dalton, 2013; Cliff et al., 2017; Vallier et al., 2009; Xu et al., 2008) have been extensively investigated as pluripotency regulators (see Introduction), while KLF7 is highly similar to Klf2/4/5 (McConnell and Yang, 2010), which are well known pluripotency regulators in mESCs and they are conserved in human (Jiang et al., 2008; Nakagawa et al., 2008). Conversely, ZNF398 is an understudied gene that was never implicated in regulation of pluripotency, thus I focused on it for subsequent experiments.

I have already observed that ZNF398 induces *PRDM14* by qPCR. To better characterize its mechanism of action, I interrogated RNA-seq data and investigated

the pluripotency factors specifically activated by ZNF398. In doing so, I confirm that *PRDM14* is regulated by ZNF398 (Figure 6D, left). Furthermore, I found that the known pluripotency factor, *LIN28B* (Zhang et al., 2016), and the novel factor *KLF7* identified in this work, were robustly activated by ZNF398 (Figure 6D).

In sum, I conclude that ZNF398 maintains pluripotency positively activating other pluripotency regulators such as PRDM14, LIN28B and KLF7.

ZNF398 represses hPSCs differentiation and sustains epithelial character

When TGF-beta is blocked hPSCs lose pluripotency, thus they differentiate and undergo a morphological change (Figure 2). After focusing on the pluripotency regulators (Figure 4-6), I decided to study the global effect of TGF-beta on hPSCs function. Thus, I performed an unbiased transcriptional analysis and observed that upon SB43 treatment 538 genes were downregulated (as already shown in Figure 6C) and 717 were upregulated (Figure 7A, in blue and yellow, respectively). In order to understand what are the biological processes controlled by these differentially expressed genes, I interrogated the DAVID (Database for Annotation, Visualization and Integrated Discovery) database. Gene Ontology 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 (Figure 7B, highlighted in red). Among them I identified a subset of genes specifically associated with epithelial character, that were downregulated by SB43 (Figure 7C, “epithelial”, left). Moreover, I observed several gene categories associated with neuroectoderm differentiation, formation and function of neural cells (Figure 7B, highlighted in green), corresponding to a set of genes upregulated by SB43 (Figure 7C, “neuroectodermal”, left). Upregulation of neuroectodermal genes was expected from studies performed in different model systems showing that TGF-beta blocks neuroectoderm formation (Muñoz-Sanjuán

and Brivanlou, 2002; Schier, 2003; Vallier et al., 2004). Indeed, inhibition of TGF-beta is conventionally used for neuroectodermal differentiation protocols (Chambers et al., 2009).
These results confirm that TGF-beta maintain pluripotency and epithelial identity and it repress neuroectodermal differentiation.

Next, I asked whether the forced expression of ZNF398 was able to counteract such transcriptional changes. Thus, I consulted the RNA-seq data and I observed that the overexpression of ZNF398 reduced the expression of neuroectodermal genes and boosted expression of epithelial genes (Figure 7C). I conclude that ZNF398 is activated by TGF-beta to maintain the correct expression of neuroectodermal and epithelial genes in hPSCs.

The experiments presented so far were all based on the forced expression of ZNF398, which showed that ZNF398 is sufficient to promote pluripotency and epithelial character.
I then asked whether ZNF398 would be required to control TGF-beta-dependent transcriptional programs. To do so, I performed siRNA-mediated knockdown of *ZNF398* and, in parallel, I also carried out knockdown of *NANOG*, as a control. First of all, I validated the knockdown efficiency induced by 5 different sequences of siRNAs and, for the following experiments, I used the 2 siRNAs showing the best knockdown effect as pool (Figure 7D, pool of siRNAs in grey). Then, I checked if the knockdown of *ZNF398* could influence self-renewal, but I observed no effect (data not shown). This is probably due to the presence of 4 factors that are individually able to maintain pluripotency downstream of TGF-beta. So, I decided to perform the knockdown experiments during the early phases (5 days of SB43) of differentiation, to see if inactivation of ZNF398 would accelerate differentiation.
In presence of SB43, *ZNF398* knockdown resulted in further reduction of pluripotency and epithelial markers, and enhanced induction of neuroectodermal genes (Figure 7E) to an extent comparable or greater than *NANOG* knockdown.

Finally, to further investigate the capacity of ZNF398 to regulate pluripotency and epithelial character of hPSCs with an independent assay, I performed Embryoid Bodies (EBs) differentiation (Figure 8A, see also Introduction and Methods for details). EBs are three-dimensional aggregates of pluripotent stem cells that are induced to differentiate into all three developmental lineages (ectoderm, mesoderm and endoderm) (Höpfl et al., 2004). This is a suitable assay to assess if the overexpression of ZNF398 could moderate the differentiation process. Actually, the forced expression of ZNF398 was able to activate expression of pluripotency and epithelial markers relative to control cells (Figure 8B, blue and orange bars, respectively). Instead, mesenchymal markers were downregulated by the overexpression of ZNF398 (Figure 8B, yellow bars). Moreover, markers of the 3 germ layers, with the exception of *T/BRA* (*BRACHYURY*), were repressed by ZNF398 (Figure 8B, green bars).

Collectively, these results indicate that ZNF398 promotes the expression of pluripotency and epithelial markers and represses genes associated with differentiation of hPSCs.

ZNF398 acts as a transcriptional activator in concert with SMAD3 and EP300

In order to investigate the molecular mechanism by which ZNF398 promotes pluripotency and epithelial character, I performed ChIP-seq for ZNF398 in 2 different hESC lines (H9 and BG01V). ZNF398 contains several zinc-finger domains and it has been shown to recognise specific DNA sequences in COS-1 cells (Conroy et al., 2002). Motif enrichment analysis at genomic regions bound by ZNF398 identified a motif similar to other ZNF factors (Figure 9A; Imbeault et al., 2017).

Cooperative binding among transcription factors has been reported in several stem cell systems (Chronis et al., 2017; Ferraris et al., 2011), thus I asked how similar the genome-wide binding profile of ZNF398 is to those of other transcriptional regulators. To do so, I interrogated available ChIP-seq data for core pluripotency

factors (NANOG, OCT4 and SOX2), regulators of Polycomb complex (EZH2 and SUZ12), general transcriptional regulators (RBBP5, PHF8, POLR2A and SAP30), regulators of chromatin architecture (SMC1, SMC3 and CTCF), SMAD3 and for the histone acetyltransferase EP300 (EP300) using CODEX database (Sánchez-Castillo et al., 2015). Surprisingly, I found that ZNF398 clustered more closely with SMAD3 and EP300, compared to the core pluripotency factors OCT4/NANOG/SOX2 or the Polycomb components (Figure 9B, top panel). Then, I conducted a similar analysis on histone modifications. So, I interrogated available ChiP-seq data for histone modifications associated with repressive or active chromatin and I observed that the peaks of ZNF398 correlate with regions decorated by acetylation of histone 3 on lysine 9 or 27 (Figure 9B, bottom panel). Such results indicate that ZNF398 co-localises with SMAD3, the histone acetyltransferase EP300 and acetylated histones.

Clustering results are confirmed by strong colocalization of ZNF398, SMAD3, EP300 and H3K27ac (Figure 9C). Histone acetylation is associated with both active promoters and enhancers, so I looked at the distribution of mono- and tri-methylation of histone 3 on lysine 4, associated with active enhancers and promoters, respectively. I looked at ZNF398 peaks and found that 3,595 ZNF398 peaks out of 5,771 appeared as active enhancers (high levels of H3K4me1 and low H3K4me3), while the remaining 2,176 peaks as active promoters (high levels of H3K4me3 and low H3K4me1).

I conclude that ZNF398 preferentially colocalises with SMAD3 and EP300 at active promoters and enhancers in hPSCs rather than with the canonical pluripotency factors OCT4 and NANOG.

In figure 9D I am presenting some snapshots of ChiP-seq data of ZNF398 (2 replicates, H9 and BG01V), EP300, SMAD3 and H3K27ac obtained with IGV (Integrative Genomics Viewer) tool. ZNF398 binds the pluripotency factor LIN28B and the epithelial master regulator ESRP1 (epithelial splicing regulatory protein 1; Warzecha et al., 2010) in collaboration with EP300 and SMAD3. Interestingly, also LEFTY1, the known TGF-beta target (Figure 3B and D), was bound by ZNF398 (Figure 9D). RNA-seq data from control hPSCs and ZNF398-expressing hPSCs showed that ZNF398

activated *LIN28B* and ESRP1, confirming the pro-pluripotency and pro-epithelial activity of ZNF398 (Figure 9D, red bar charts, left and centre).

Moreover, also *LEFTY1* was activated by ZNF398 (Figure 9D, red bar chart on the right). Similarly, other known TGF-beta direct targets were strongly induced by ZNF398 such as *LEFTY1/2*, *CER1*, *TGFB1* and *NODAL* (Figure 10A). This observation, together with the colocalization of ZNF398 with SMAD3, led me to hypothesise that ZNF398 might function as a booster of TGF-beta signaling in hPSCs by binding SMAD3 targets.

To test this hypothesis, I analysed all 81 SMAD3 direct target genes (Figure 3A) and observed that they were significantly upregulated in cells expressing ZNF398 (Figure 10B), further indicating a functional role of ZNF398 as activator of SMAD3 co-bound targets. This activity is specific of ZNF398, given that NANOG expression had no discernible effect on SMAD3-ZNF398 targets.

To further confirm these results with an independent technique in two hPSC lines, I compared PSCs expressing ZNF398 against control PSCs by measuring the expression of *LEFTY1* with qPCR. ZNF398 boosted by >10 fold the basal expression of *LEFTY1* (i.e. in the presence of TGF-beta), and was even able to maintain residual *LEFTY1* expression also in the absence of TGF-beta signaling (Figure 10C).

Finally, to obtain further insights into the molecular mechanism of action of ZNF398 on SMAD3 transcriptional activity, I quantified the cellular localization of SMAD2/3 proteins in hPSCs with immunofluorescence for total SMAD2/3. Actually, the receptor SMADs, SMAD2/3, have the ability to shuttle in and out of the nucleus under basal conditions (Massagué et al., 2005). Upon TGF-beta stimulation, SMAD2/3 form a complex with SMAD4, the co-SMAD, that translocate into the nucleus to regulate gene expression (Massagué et al., 2005; see also Introduction). Considering that ZNF398 seemed to activate TGF-beta targets, I wondered if ZNF398 could stimulate the SMAD2/3 translocation into the nucleus upon TGF-beta induction. As expected, inhibition of TGF-beta signal resulted in exclusion from the nucleus, which was

restored after 60 minutes of restimulation; on the other hand, ZNF398 accelerated and increased nuclear SMAD2/3 localisation after restimulation (Figure 10D, see also Figure 10E for the entire fields of Figure 10D).

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

ZNF398 is required for somatic cell reprogramming

So far, the results indicate that ZNF398 promotes pluripotency and epithelial character programs in hPSCs. In order to further test the function of ZNF398, I decided to investigate its role during reprogramming of somatic cells into pluripotent stem cells. Somatic cell reprogramming relies on somatic cell transfection or transduction of a specific cocktail of transcription factors in the presence of the correct cellular signals (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; see also Introduction). Specifically, it has been shown that human fibroblasts are efficiently reprogrammed into iPSCs by using the OSKMNL (OCT4, SOX2, KLF4, MYC, NANOG and LIN28A) cocktail of factors (Luni et al., 2016; Yu et al., 2007). Reprogramming from somatic cells, such as fibroblasts, requires an early Mesenchymal to Epithelial Transition (MET) followed by the activation of endogenous pluripotency factors (Høffding and Hyttel, 2015; Pei et al., 2019).

First of all, I noticed that *ZNF398* is robustly expressed in human fibroblasts (Figure 11E) raising the possibility that ZNF398 promotes acquisition of epithelial character and pluripotency during early stages of reprogramming. To test the requirement of endogenous ZNF398 for reprogramming, I reprogrammed human fibroblasts by delivery of mRNAs encoding for OSKMNL for 8 days, in combination with siRNAs and I then analysed them at day 6 and 14 for morphology and qPCR and for scoring of pluripotent colonies (Figure 11A).

First of all, I wanted to make sure that siRNA-mediated knockdown was efficient under the conditions I used. I transfected human fibroblasts with an mRNA encoding for GFP in combination with a siCONTROL or a siGFP for 3 days. In this way I could easily check whether as siRNA targeting an mRNA of interest (GFP, in this case) was

able to efficiently knock it down. In Figure 11B, I am showing the effect of siCONTROL or siGFP with a fluorescence microscope. The majority of fibroblasts transfected with GFP mRNA + siCONTROL expressed GFP protein, while GFP signal was strongly reduced in fibroblasts transfected with GFP mRNA + siGFP. So, I concluded that this experimental setup allowed efficient siRNA-mediated knockdown.

Then, I performed the reprogramming experiment transfecting mRNAs for 8 days in combination with siRNAs targeting ZNF398 or siCONTROL, or without siRNAs as an additional control (Figure 11A). By day 6 of reprogramming, fibroblasts transfected with control siRNAs formed clusters of epithelial cells, indicative of MET. Importantly, siZNF398 did not affect overtly cell viability or proliferation, given that a comparable number of cells were observed in all conditions. However, MET was specifically reduced upon ZNF398 knockdown (Figure 11C, see in particular close up) in line with ZNF398 acting as an inducer of epithelial character.

Around day 10 small colonies emerged and were stabilised over the following 4 days. In the normal conditions (no siRNA and siCONTROL), the number of colonies obtained per channel was about 7. Conversely, *ZNF398* knockdown resulted in a 6.36-fold reduction in the number of iPSC colonies, relative to Control siRNA (Figure 11D).

Finally, I compared siCONTROL and siZNF398 at day 6 and day 14 to fibroblasts and to KiPS cell lines measuring the expression of pluripotency and epithelial markers using qPCR. First, I confirmed the effective reduction by ~50% of ZNF398 at day 6 by siRNAs (Figure 11E, top panel of the left). I then observed that *ZNF398* knockdown failed to activate a large panel of pluripotency and epithelial markers at day 14 (Figure 11E and Figure 11F).

Taken together, these results show the requirement of ZNF398 for efficient generation of iPSCs from fibroblasts.

ZNF398 promotes 3D self-organization

After implantation, epiblast cells self-organize in a columnar epithelium that undergoes lumenogenesis (Deglincerti et al., 2016; Shahbazi et al., 2016; see also Introduction), while maintaining pluripotency. It is possible to recapitulate such processes *in vitro* by surrounding hPSCs in extracellular matrix together with media supporting pluripotency (Shahbazi et al., 2016; Shao et al., 2016, 2017; Taniguchi et al., 2015; see also Introduction). In all these different published protocols (Shahbazi et al., 2016; Shao et al., 2016, 2017; Taniguchi et al., 2015), I noticed that TGF-beta ligands are consistently present. So, I asked whether pluripotency and morphological organization could be coupled and under the control of TGF-beta. To test this hypothesis, I plated a suspension of single hPSCs on top of layer of Matrigel in a medium with TGFB1 and 5% of Matrigel (Figure 12A) (adapted from Shahbazi et al., 2016). In presence of TGB1, I observed efficient formation of cell aggregates that after 2 days started to form a lumen (data not shown). At day 4, I checked the morphology and I performed an immunofluorescence for the pluripotent marker NANOG and for cytoskeletal and epithelial markers F-ACTIN (Phalloidin) and BETA-CATENIN. I observed that these cell aggregates grew as large spheroids composed by a columnar epithelium (Figure 12B and C, top left panel). Moreover, transcriptional analysis performed by qPCR revealed that such self-organized structures expressed epithelial and pluripotency markers at levels comparable to (e.g. *ECAD* and *LIN28A*) or markedly higher than (*EPCAM* and *NANOG*, *PRDM14*, *UTF1*, *ZNF398*) parental hPSCs cultured under conventional conditions (Figure 12F) indicating the maintenance of pluripotency and epithelial identity in 3D culture condition.

Strikingly, in the presence of SB43, I observed formation of cellular aggregates with slightly lower efficiency (SB43 93.67+/-13.92 vs DMSO 128.33+/-25.39), however cells appeared disorganised and elongated (Figure 12C, top panel and Figure 12D). Furthermore, several pluripotency and epithelial markers were downregulated by SB43 while mesenchymal markers were strongly upregulated (Figure 12E and 12F).

I conclude that TGF-beta signal is critical for self-organization of hPSCs *in vitro*.

Moreover, I noticed that *ZNF398* was upregulated during self-organization and strongly downregulated by SB43 (Figure 12E and 12F). Given its ability to induce pluripotency and epithelial genes, I thought that it could play a role also during self-organization of hPSCs. To test this hypothesis, I generated hPSCs expressing ZNF398 at ~2-fold the endogenous levels (Figure 12E) and plated them in presence or absence of SB43 under conditions allowing self-organization. I observed formation of organized epithelial spheroids even in the presence of SB43. Epithelial markers were highly induced, mesenchymal markers were ablated (Figure 12E and 12F), while half of the pluripotency markers tested were maintained by ZNF398 forced expression.
I conclude that ZNF398 is epistatic to TGF-beta signal for self-organization of hPSCs.

DISCUSSION

TGF-beta signal is critical for hPSC self-renewal (Vallier et al., 2009; Xu et al., 2008). The transcription factor Nanog was first identified in murine ESCs for its capacity to maintain pluripotency in the absence of exogenous signals (Chambers et al., 2003). The important role of NANOG in the maintenance of pluripotency was found conserved also in hPSCs (Darr et al., 2006). It was then shown that TGF-beta directly induces NANOG expression in hPSCs (Vallier et al., 2009; Xu et al., 2008). However, an unbiased and systematic analysis of TGF-beta functional mediators in hPSCs was still missing. For this reason, I performed a transcriptome-level analysis of TGF-beta targets followed by a gain-of-function screening to identify novel pluripotency regulators.

Loss-of-function screenings have been performed in hPSCs, whereby genes were inactivated by RNA interference or using the CRISPR systems (Chia et al., 2010; Yilmaz et al., 2018). Such studies identified novel pluripotency regulators, such as PRDM14 (Chia et al., 2010). However, loss-of-function approaches might fail to identify critical regulators because of the functional redundancy with other factors. For example, a CRISPR screening in murine ESCs failed to identify the majority of known pluripotency factors (Hackett et al., 2018), likely because the pluripotency network is highly redundant and robust to inactivation of single factors (Dunn et al., 2014). For these reasons, I 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. This approach allowed the identification of several critical murine pluripotency regulators (Martello and Smith, 2014).

I identified a quartet of transcription factors, NANOG, MYC, KLF7 and ZNF398, which individually promotes hPSC self-renewal. In particular, ZNF398 is able to sustain both pluripotency and the epithelial genes downstream TGF-beta (Figure 13A). Indeed, ZNF398 promotes the expression of pluripotency and epithelial markers and represses neuroectodermal and mesenchymal markers that are associated with differentiation of hPSCs. I am currently generating *ZNF398* knockout hPSCs by

CRISPR-Cas9 system in order to further confirm its role in the maintenance of pluripotency and the epithelial character.

Of note, the conventional hPSCs are in a pluripotent state that is defined as “primed” to differentiate (Nichols e Smith, 2009). Conversely, the conventional mESCs are in a more “naïve” pluripotent state. Subsequently, mouse embryonic stem cells were also derived from the post-implantation epiblast and they are called mouse Epiblast stem cells (mEpiSCs, Brons et al., 2007; Tesar et al., 2007). These cells showed some features that are similar to hPSCs (Nichols e Smith, 2009) and they are considered the mouse counterpart of hPSCs. So, it would be nice to test if Zfp398, the orthologous gene of ZNF398, has a similar role in the maintenance of pluripotency in mEpiSCs.

Additionally, naïve hPSCs have been recently derived (Giulitti et al., 2018; Guo et al., 2016; Takashima et al., 2014) and I noticed that ZNF398 and also KLF7 are expressed also in these naïve cell lines. It would be interesting to check if the role of TGF-beta in the maintenance of pluripotency and its direct targets are conserved among primed and naïve human PSCs.

Moreover, it will be interesting to understand if the role of ZNF398 as a mediator of TGF-beta is a conserved mechanism also in somatic cell lines and in cancer cell lines. I have interrogated available transcriptomic data of different human tissues and several cancer cell lines and observed that ZNF398 is expressed at detectable levels only in a small fraction of cancer lines. It would be interesting to test in the future whether in such cells ZNF398 is playing any functional role, as previously suggested for OCT4 and NANOG that have been found expressed in some cancers (Jeter et al., 2015; Villodre et al., 2016).

ZNF398 knockdown strongly reduced reprogramming efficiency, indicating a critical role during establishment of pluripotency. Specifically, I 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 can be used in the reprogramming cocktail to generate iPSCs at higher efficiency or to identify fully reprogrammed cells.

Human PSCs cultured in 2D form flat epithelia (Figure 2A and 2B). Self-organization into spherical cysts of columnar polarised epithelium (Shahbazi et al., 2016; Shao et al., 2016, 2017; Taniguchi et al., 2015) is associated with increased expression of both epithelial and pluripotency markers (Figure 12E and 12F), suggesting that the epithelial organization might have a positive impact on the pluripotency network, potentially by mimicking the natural environment experienced by pluripotent cells of the embryo. I observed that TGF-beta inhibition leads to a reduction of epithelial markers and a concomitant increase in mesenchymal markers, both in 2D and 3D culture. ZNF398 is able to induce epithelial character and block mesenchymal markers expression (Figures 7E, 8B, 12E and 12F).

Moreover, I demonstrated that TGF-beta is fundamental for self-organization of hPSCs and ZNF398 forced expression allows self-organization of hPSCs also when TGF-beta signal is absent, indicating that ZNF398 is a critical player downstream of TGF-beta signal for the formation of 3D structures. The growing field of 3D self-organization will allow to study early events in human development *in vitro*. It will be interesting to characterise the transcriptional profiles of single hPSCs cultured in 3D and compare them to available single-cell RNAseq datasets from human embryos in order to understand if 3D culture better recapitulates molecular events observed during early embryonic development. Moreover, it will be interesting to perform the single-cell RNAseq of the ZNF398 overexpressing cells in order to better understand the molecular mechanism by which ZNF398 regulates self-organization.

Finally, it will be interesting to apply computational modelling (Dunn et al., 2014) to reconstruct the network of interactions among such factors in order to study how the human pluripotency network is maintained or dissolved during differentiation.

For example, I observed that NANOG expression is more heterogeneous than the expression of OCT4 at the level of single cells (Figure 6A and 6B) when ZNF398 is over-expressed. This clearly indicate that OCT4 and NANOG are regulated in a different

way. NANOG is a direct target of TGF-beta (Vallier et al., 2009) while OCT4 is not, and ZNF398 seems to induce specifically NANOG expression (Figure 4D).
Integrating all the information we collected about the regulation of pluripotency factors in a single coherent model requires a computational approach, which we are currently undertaking. We believe that such computational model will be useful for the study of PSC differentiation and reprogramming.

Finally, the identification of novel human pluripotency factors, such as ZNF398, will have broad implications for the generation, characterization and use of hiPSCs. A common limitation of hiPSCs is their heterogeneity and their bias in differentiation: different hPSCs express similar levels of pluripotency markers but show different propensity to differentiate (Osafune et al., 2008). Until now, the same genes used in the cocktail for reprogramming are often, and paradoxically used for the identification of hiPSCs. For this reason, novel pluripotency regulators could be used as markers of “good quality” hiPSCs or used to improve the efficiency and quality of iPS generation.

FIGURES

Figure 1. Introduction

(A) Human Pluripotent Stem Cells (hPSCs) can be derived from the epiblast of pre-implantation blastocyst as human Embryonic Stem Cells (hESCs) or from somatic cells via transcription factors-mediated reprogramming as induced pluripotent stem cells (hiPSCs). The blastocyst consists of three cell types: the trophoblast, the hypoblast and the epiblast. Only the epiblast will give rise to embryonic tissues, while trophoblast and hypoblast will form extra-embryonic support structures.
hESCs and hiPSCs can be expanded *in vitro*

(B) Pathways activated by TGF-beta/Activin. TGF-beta is fundamental for the maintenance of pluripotency and self-renewal in hPSCs (see introduction). Note the action of SB43 (SB431542), which is a known inhibitor of TGF-beta signaling that acts at the type I receptor (ALK).

(C) Diagram representing the network of transcription factors activated by TGF-beta that maintain pluripotency in hPSCs. The majority of downstream mediators of TGF-beta are still unknown.

(D) Experimental approach used to form Embryoid Bodies (EBs) from hPSCs.
Small clumps or single cells suspension of hPSCs are put in a low attachment plate in a medium that allow differentiation (DMEM+20% FBS). After some days, EBs are formed.

(E) Scheme illustrating the self-organization process *in vivo* and *in vitro* (see Introduction).

FIGURE 1

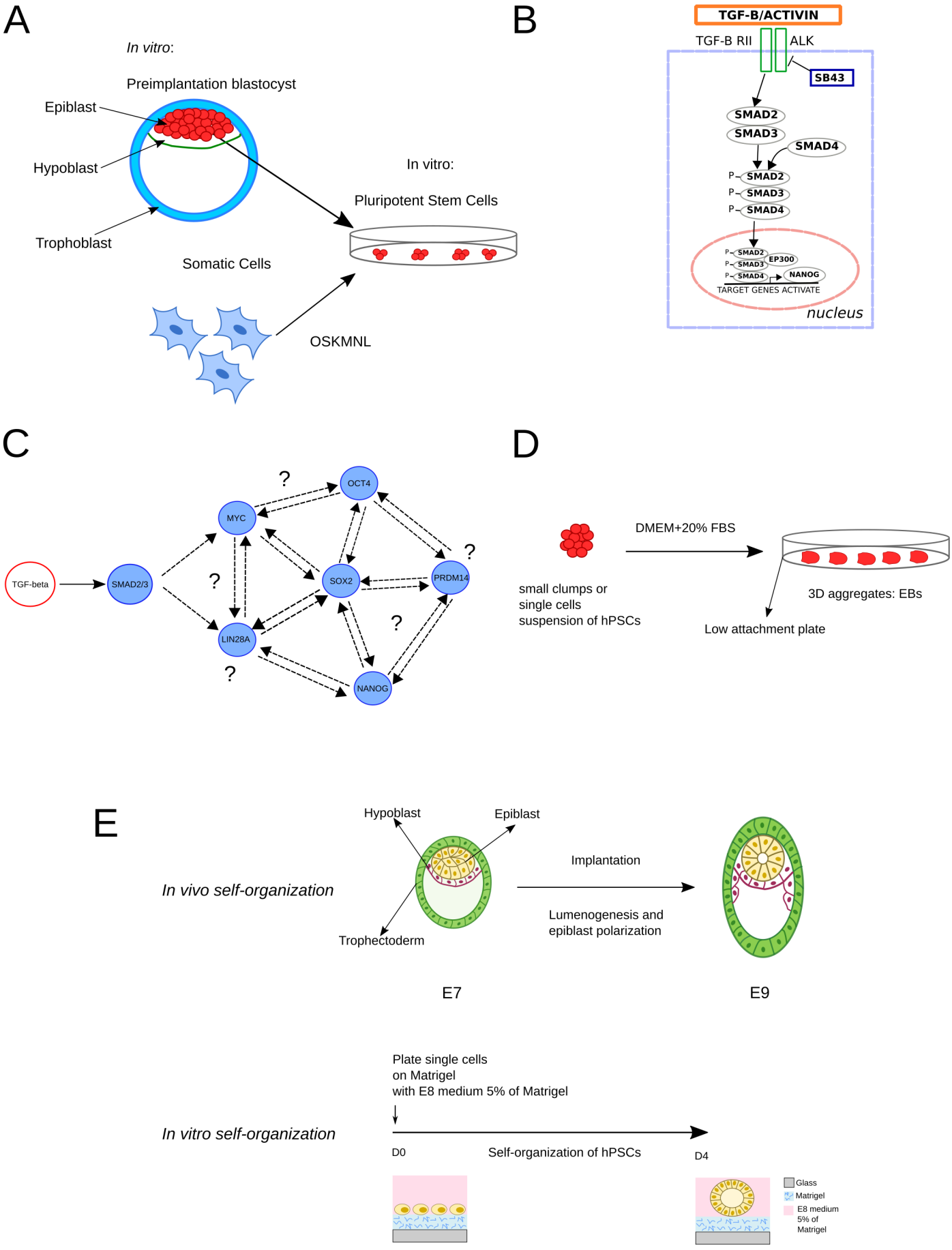

A
In vitro:
Preimplantation blastocyst
Epiblast
Hypoblast
Trophoblast
In vitro:
Pluripotent Stem Cells
Somatic Cells
OSKMNL
B
TGF-B/ACTIVIN
TGF-B RII
ALK
SB43
SMAD2
SMAD3
SMAD4
P-SMAD2
P-SMAD3
P-SMAD4
P-SMAD2
P-SMAD3
P-SMAD4
EP300
NANOG
TARGET GENES ACTIVATE
nucleus
C
TGF-beta
SMAD2/3
MYC
OCT4
SOX2
PRDM14
LIN28A
NANOG
?
D
DMEM+20% FBS
small clumps or
single cells
suspension of hPSCs
3D aggregates: EBs
Low attachment plate
E
In vivo self-organization
Hypoblast
Epiblast
Trophectoderm
Implantation
Lumenogenesis and
epiblast polarization
E7
E9
In vitro self-organization
Plate single cells
on Matrigel
with E8 medium 5% of Matrigel
D0
Self-organization of hPSCs
D4
Glass
Matrigel
E8 medium
5% of
Matrigel

Figure 2. TGF-beta is required for self-renewal and epithelial identity in hPSCs

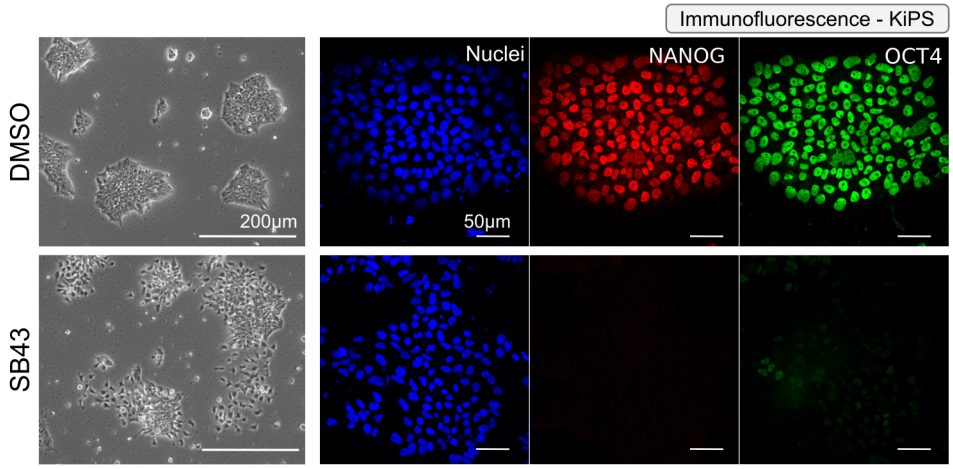
(A) Left: Morphology of human Keratinocytes-derived iPSCs (KiPS) treated with TGF-beta receptor inhibitor SB431542 (SB43, 10µM) or the vehicle DMSO for 5 days. Right: Immunostaining for the pluripotency markers NANOG and POU5F1/OCT4 shows a reduction of both markers after 5 days of SB43 treatment. Nuclei were identified by DAPI (4',6-diamidino-2-phenylindole) staining.

(B) Immunostaining for E-CADHERIN (*ECAD*, *CDH1*) shows a reduction of expression after 5 days of SB43 treatment, while immunostaining for VIMENTIN (*VIM*) shows an increased expression. Nuclei were identified by DAPI staining.

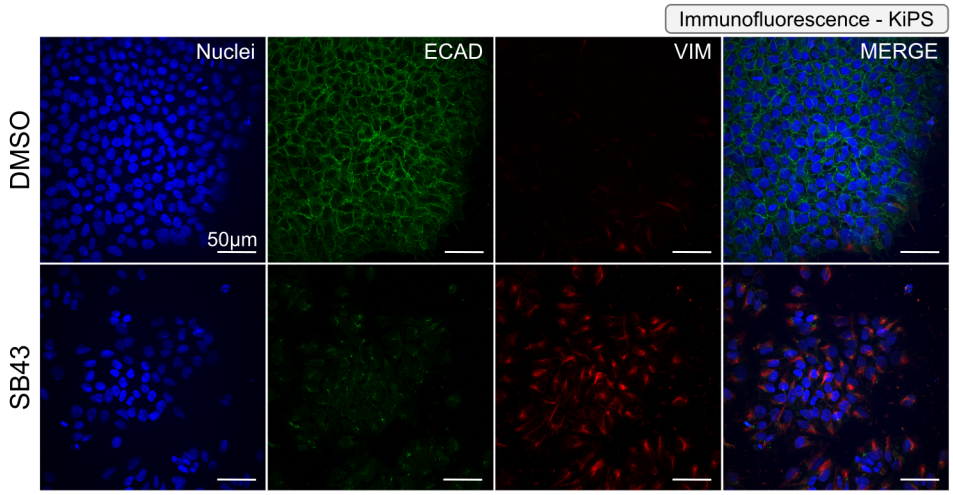
(C) Gene-expression analysis by qPCR of KiPS treated with SB43 or DMSO for 5 days. Bars indicate mean and s.e.m. of 4 independent experiments shown as dots. Expression was normalised to the mean of DMSO samples. Unpaired two-tailed Mann–Whitney *U* test.

FIGURE 2

A


Immunofluorescence - KiPS
DMSO
SB43
200µm
Nuclei
NANOG
OCT4
50µm

B


Immunofluorescence - KiPS
DMSO
SB43
Nuclei
ECAD
VIM
MERGE
50µm

C

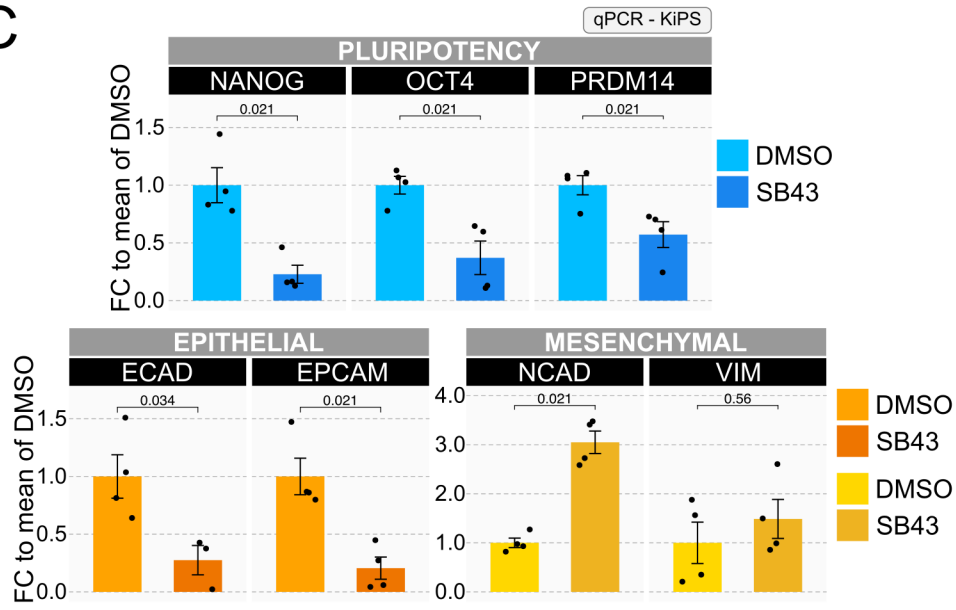

qPCR - KiPS
PLURIPOTENCY
NANOG
OCT4
PRDM14
0.021
0.021
0.021
FC to mean of DMSO
1.5
1.0
0.5
0.0
DMSO
SB43
EPITHELIAL
ECAD
EPCAM
0.034
0.021
MESENCHYMAL
NCAD
VIM
0.021
0.56
4.0
3.0
2.0
1.0
0.0
DMSO
SB43
DMSO
SB43

Figure 3. Identification of TGF-beta transcriptional targets in hPSCs

(A) Approach used to identify potential SMAD3 direct targets.

(B) Top: Transcriptome analysis of hESCs treated with SB43 for 48 hours (Microarray data from Vallier et al., 2009). Dark grey dots indicate differentially Expressed Genes (DEGs) for -1 > Log2 fold-change > 1 and p-value < 0.05. Orange dots refer to DEGs bound by SMAD3 (data from Mullen et al., 2011), max distance between peak midpoint and TSS: +/- 50kb). Bottom: Transcriptome analysis of hiPSCs treated for 4 hours with mTeSR after 16 hours of SB43 treatment (RNA-seq data, in this study). Dark grey dots indicate DEGs for -0.585 > Log2 fold-change > 0.585 (corresponding to an increase of 50%) and p-value < 0.05. Orange dots refer to DEGs bound by SMAD3. Known SMAD3 targets, such as NANOG, LEFTY2, SKIL and SMAD7 serve as positive controls (Shi and Massagué, 2003; Vallier et al., 2009; Xu et al., 2008).

(C) Bar plot showing RNA-seq data of H9 (WA09 wicell) (data from Rada-Iglesias et al., 2011, GSE24447, GSM602289). Bars indicate absolute expression (RPKM) of SMAD3-bound and differentially expressed Transcription Factors (TFs). Dashed line indicates the chosen threshold (>3RPKM) and dark orange bars highlight the TFs above it.

(D) Balloon plot summarising gene-expression analysis for the indicated SMAD3 target genes. The size of each balloon indicates the statistical significance (Unpaired two-tailed *t*-test), while the colour indicates the fold-change in expression. Left: Microarray data refer to the experiment of Figure 3B top panel, which were independently validated by qPCR. Validation experiments were performed in two different culture conditions: feeder-free or on feeders (MEF). Expression was normalised to DMSO treated cells. Right: RNA-seq data refer to the experiment of Figure 3B bottom panel, which were independently validated by qPCR. Expression was normalized to SB43 treated samples. Validation experiments by qPCR were performed both under feeder-free conditions, with TGF-beta induction (2ng/ml) for

1 hour or for 4 hours or on feeders with TGF-beta induction for 1 hour. Additionally, Activin A stimulation (25ng/ml) on feeders was performed for 1 hour. 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.

(E) Example of SMAD3 binding and gene expression analysis of three validated targets (SMAD7, NANOG and ZNF398). Left: Gene tracks represent binding of SMAD3 (data from Mullen et al., 2011) at the indicated gene loci. Right: Gene expression analysis by qPCR. See also Figure 3D. Bars indicate mean and s.e.m. of independent experiments shown as dots. Expression was normalised to the mean of SB43 samples. Unpaired two-tailed *t*-test: * $P<0.05$; ** $P<0.01$.

FIGURE 3

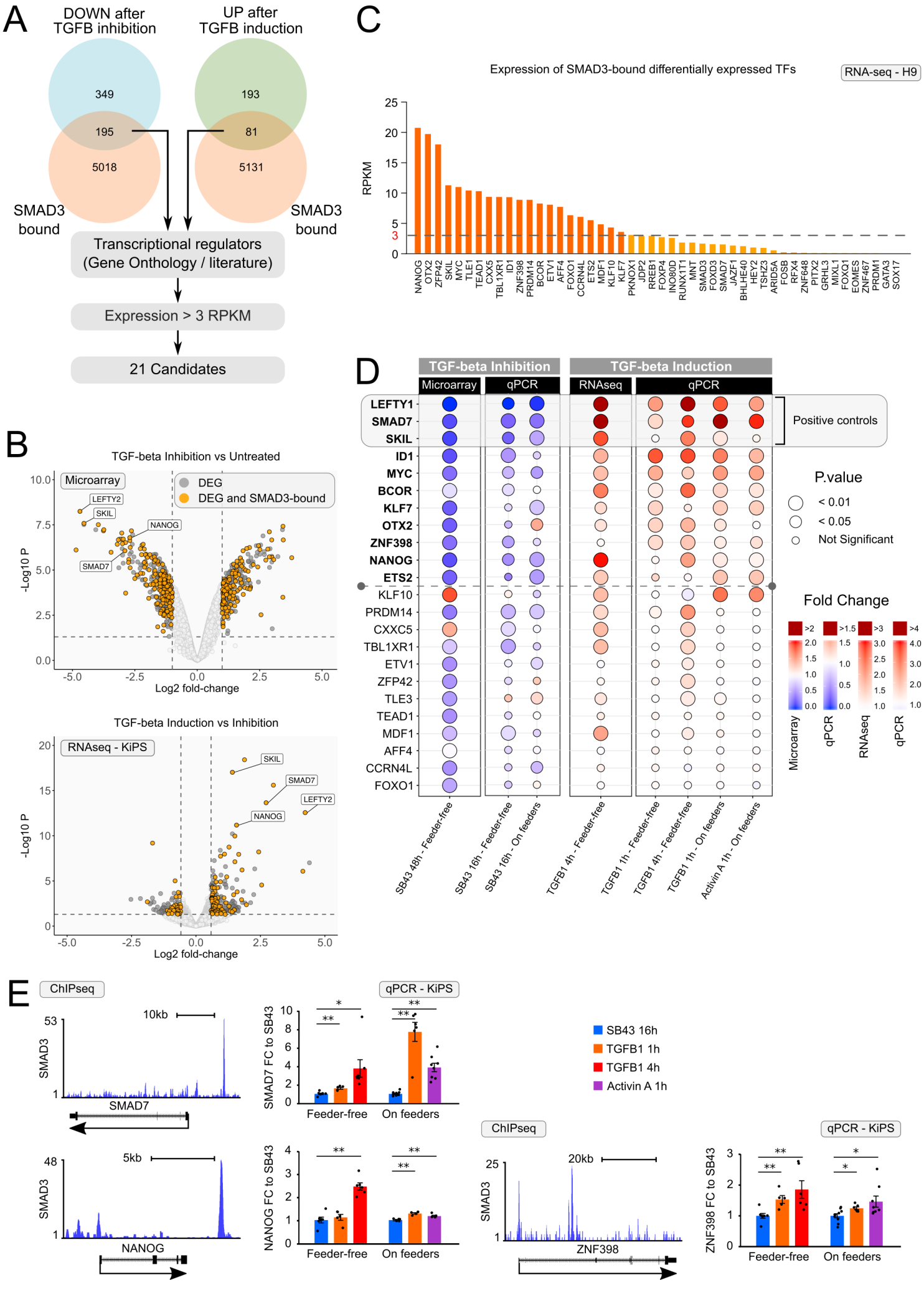

A
DOWN after TGFB inhibition
UP after TGFB induction
349
195
5018
193
81
5131
SMAD3 bound
SMAD3 bound
Transcriptional regulators (Gene Onthology / literature)
Expression > 3 RPKM
21 Candidates
B
TGF-beta Inhibition vs Untreated
Microarray
DEG
DEG and SMAD3-bound
LEFTY2
SKIL
NANOG
SMAD7
-Log10 P
Log2 fold-change
TGF-beta Induction vs Inhibition
RNAseq - KiPS
SKIL
SMAD7
LEFTY2
NANOG
-Log10 P
Log2 fold-change
C
Expression of SMAD3-bound differentially expressed TFs
RNA-seq - H9
RPKM
NANOG OTX2 ZFP42 SKIL MYC TLE1 TEAD1 CXX5 TBL1XR1 ID1 ZNF398 PRDM14 BCOR ETV1 AFF4 FOXO1 CCRN4L ETS2 MDF1 KLF10 KLF7 PKNOX1 JDP2 RREB1 FOXP4 INO80D RUNX1T1 MNT SMAD3 FOXD3 SMAD7 JAZF1 BHLHE40 HEY2 TSHZ3 ARID5A FOSB RFX4 ZNF648 PITX2 GRHL3 MIXL1 FOXQ1 EOMES ZNF467 PRDM1 GATA3 SOX17
D
TGF-beta Inhibition
TGF-beta Induction
Microarray
qPCR
RNAseq
qPCR
LEFTY1
SMAD7
SKIL
Positive controls
ID1
MYC
BCOR
KLF7
OTX2
ZNF398
NANOG
ETS2
KLF10
PRDM14
CXXC5
TBL1XR1
ETV1
ZFP42
TLE3
TEAD1
MDF1
AFF4
CCRN4L
FOXO1
P.value
< 0.01
< 0.05
Not Significant
Fold Change
Microarray
qPCR
RNAseq
qPCR
SB43 48h - Feeder-free
SB43 16h - Feeder-free
SB43 16h - On feeders
TGFB1 4h - Feeder-free
TGFB1 1h - Feeder-free
TGFB1 4h - Feeder-free
TGFB1 1h - On feeders
Activin A 1h - On feeders
E
ChIPseq
10kb
SMAD3
SMAD7
qPCR - KiPS
SMAD7 FC to SB43
Feeder-free
On feeders
SB43 16h
TGFB1 1h
TGFB1 4h
Activin A 1h
5kb
NANOG
NANOG FC to SB43
ChIPseq
20kb
ZNF398
qPCR - KiPS
ZNF398 FC to SB43

Figure 4. Functional identification of pluripotency regulators in hPSCs (KiPS and HES2 cell lines)

(A) Experimental approach used to test the functional role of SMAD3 targets identified in Figure 3. hPSCs were co-transfected with pBase helper plasmid (encoding for the transposase) and a piggyBac vector allowing genomic integration and the stable expression of candidate SMAD3 targets or an empty vector (Empty). After 48 hours, hygromycin selection (200µg/ml) was added for at least 2 weeks. After selection, cells were cultured with or without SB43 for 5 days and analysed as indicated.

(B) Left: Clonal assay quantification of hESCs (HES2) (red bars) and hiPSCs (KiPS) (orange bars) stably expressing an empty vector control (Empty) or 8 different SMAD3 targets identified in Figure 3D. 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 average percentage of AP positive colonies. Dots represent independent experiments. Unpaired two-tailed Mann–Whitney *U* test relative to Empty SB43 samples. Right: Representative AP staining images of clonal assay performed in KiPS.

(C) Representative morphology of HES2 colonies stably expressing an empty vector (Empty) in presence of DMSO or SB43 and HES2 stably expressing the 8 SMAD3 targets in presence of SB43.

(D) Gene expression analysis by qPCR of HES2 (light green bars) and KiPS (dark green bars) stably expressing an Empty vector or the 8 SMAD3 targets and treated with or without SB43 for 5 days. Bars indicate mean and s.e.m of independent experiments, shown as dots. Expression was normalised to the Empty DMSO samples. Unpaired two-tailed Mann–Whitney *U* test.

FIGURE 4

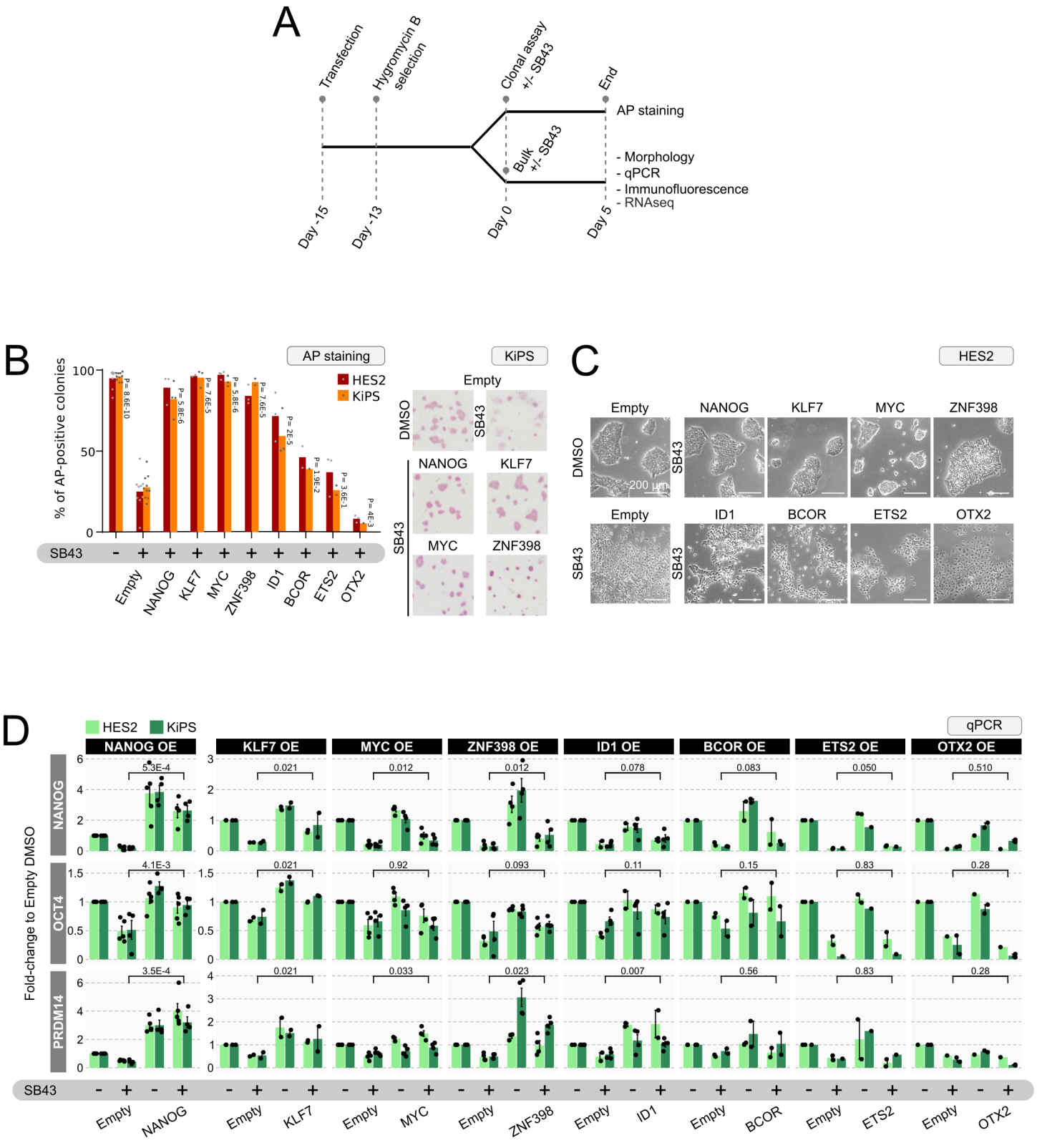

A
Transfection
Hygromycin B selection
Clonal assay +/- SB43
End
AP staining
Bulk +/- SB43
- Morphology
- qPCR
- Immunofluorescence
- RNAseq
Day -15
Day -13
Day 0
Day 5
B
AP staining
KiPS
% of AP-positive colonies
HES2
KiPS
SB43
Empty
NANOG
KLF7
MYC
ZNF398
ID1
BCOR
ETS2
OTX2
DMSO
C
HES2
200 µm
D
HES2
KiPS
qPCR
NANOG OE
KLF7 OE
MYC OE
ZNF398 OE
ID1 OE
BCOR OE
ETS2 OE
OTX2 OE
Fold-change to Empty DMSO
NANOG
OCT4
PRDM14

Figure 5. Functional identification of pluripotency regulators in H9 cell line

(A) Left: Clonal assay quantification of H9 stably expressing an empty vector, NANOG, KLF7, MYC or ZNF398. Five 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 average percentage of AP positive colonies. Dots represent independent experiments. Unpaired two-tailed Mann–Whitney *U* test. Right: Representative AP staining images of clonal assay performed with H9.

(B) Representative morphology of H9 colonies stably expressing empty vector (Empty) in presence of DMSO or SB43 and H9 stably expressing NANOG, KLF7, MYC or ZNF398 in presence of SB43 for 5 days.

(C) Top: Evaluation of overexpression levels by qPCR for H9, HES2 and KiPS (red, green and blue bars respectively) stably expressing an empty vector, NANOG, KLF7, MYC or ZNF398 and treated with SB43 or DMSO for 5 days. Bars indicate mean and s.e.m. of independent experiments shown as dots. Expression was normalised to Empty DMSO sample. We observe consistently reduced transgene expression for MYC and ZNF398 in H9 cells which might explain the reduced phenotypic effect. Bottom: Measurement of transgene expression levels by qPCR. Primers were designed to recognise a portion of the 3'UTR shared by all transgenes, allowing direct comparison among different vectors and specific detection of exogenous transcripts. Data for H9, HES2 and KiPS (red, green and blue bars respectively) stably expressing an empty vector, NANOG, KLF7, MYC or ZNF398 and treated with DMSO for 5 days are shown. Bars indicate mean and s.e.m. of independent experiments shown as dots. Expression was normalised to H9 Empty samples. Untransfected KiPS samples serve as negative control. Exogenous expression was comparable among the 4 candidates; however, it was consistently lower in H9 cells, especially for MYC and ZNF398.

FIGURE 5

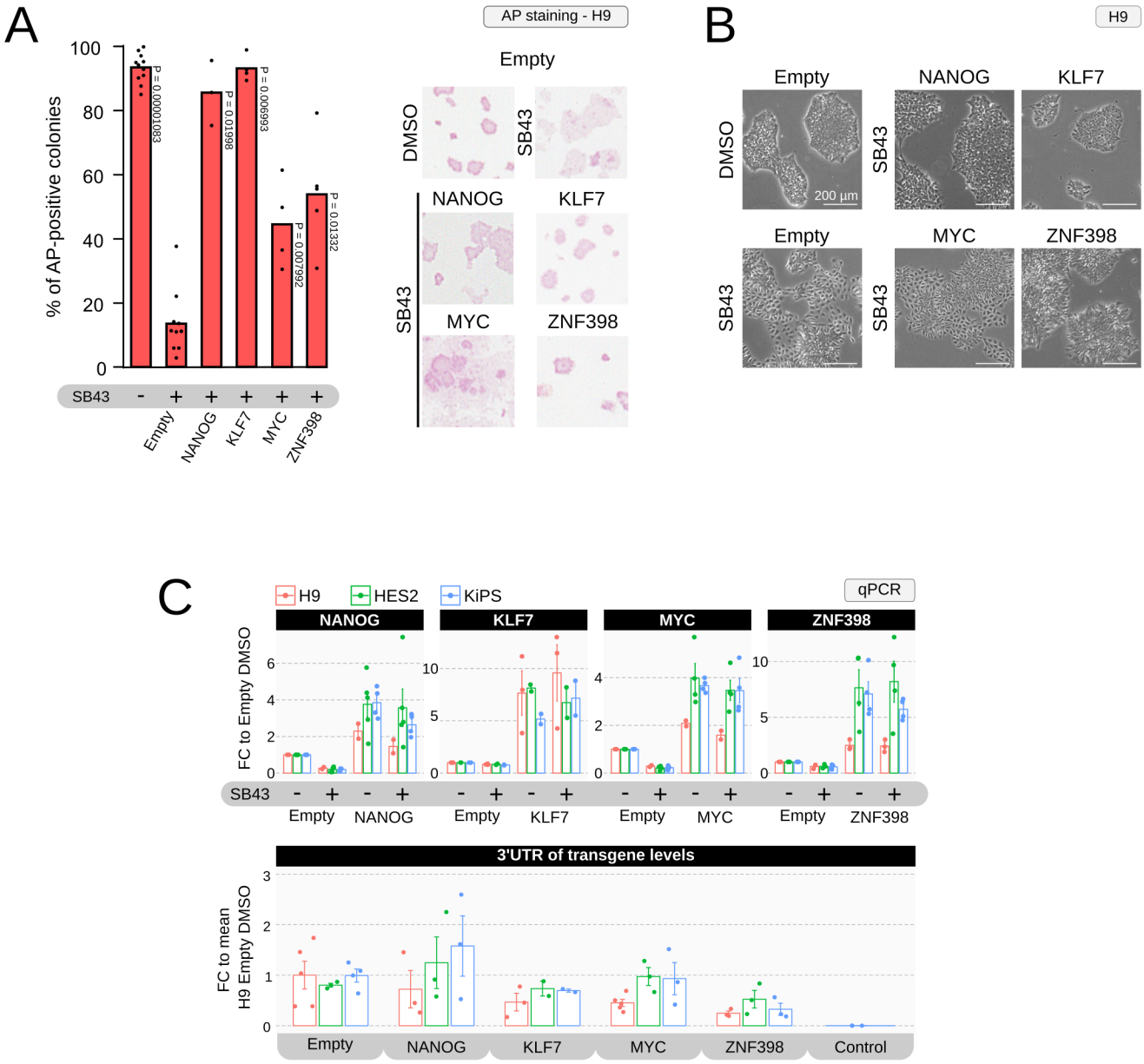

A
AP staining - H9
% of AP-positive colonies
100
80
60
40
20
0
P = 0.0000108
P = 0.01398
P = 0.006993
P = 0.007992
P = 0.01332
SB43 - + + + + +
Empty
NANOG
KLF7
MYC
ZNF398
Empty
DMSO
SB43
NANOG
KLF7
SB43
MYC
ZNF398
B
H9
Empty
NANOG
KLF7
DMSO
SB43
200 µm
Empty
MYC
ZNF398
SB43
SB43
C
H9
HES2
KiPS
qPCR
NANOG
KLF7
MYC
ZNF398
FC to Empty DMSO
SB43 - + - + - + - + - + - + - + - +
Empty
NANOG
Empty
KLF7
Empty
MYC
Empty
ZNF398
3'UTR of transgene levels
FC to mean H9 Empty DMSO
Empty
NANOG
KLF7
MYC
ZNF398
Control

Figure 6. Functional identification of pluripotency regulators in KiPS and H9 with immunofluorescence and RNA-seq

(A) Left: representative images of immunostaining for the pluripotency markers NANOG and POU5F1/OCT4 of KiPS stably expressing an empty vector control (Empty) in presence of DMSO or SB43 and KiPS stably expressing NANOG, KLF7, MYC or ZNF398 in presence of SB43 for 5 days. Nuclei were identified by DAPI staining. Right: Violin plots showing fluorescence intensity quantification of NANOG and OCT4. For each condition, at least 300 nuclei from 5 randomly selected fields were analysed.

(B) Left: representative images of immunostaining for the pluripotency markers NANOG and POU5F1/OCT4 of H9 transfected with an empty vector (Empty) in presence of DMSO or SB43 and H9 stably expressing NANOG, KLF7, MYC or ZNF398 in presence of SB43 for 5 days. Nuclei were identified by DAPI staining. Right: Violin plots showing fluorescence intensity quantification of NANOG and OCT4. For each condition, at least 300 nuclei over 5 images were analysed.

(C) Box plot showing absolute expression levels (normalised counts, TPM) of 538 genes DOWN-regulated by SB43 treatment (5 days) in KiPS stably expressing an empty vector (see Figure 7A, blue dots). Shown data refers to KiPS transfected with the empty vector in presence of DMSO or SB43 and for KiPS stably expressing NANOG, KLF7, MYC or ZNF398 in presence of SB43 for 5 days. Average fold-change relative to Empty SB43 sample (X) is reported for each condition. Unpaired two-tailed *t*-test: **** $P<0.0001$.

(D) Bar charts showing gene expression analysis by RNA-seq of KiPS stably expressing an empty vector (Empty) or ZNF398 in presence of DMSO or SB43 for 5 days. Bars indicate the mean and the s.e.m of at least 2 independent experiments. Absolute expression is reported as TPM.

FIGURE 6

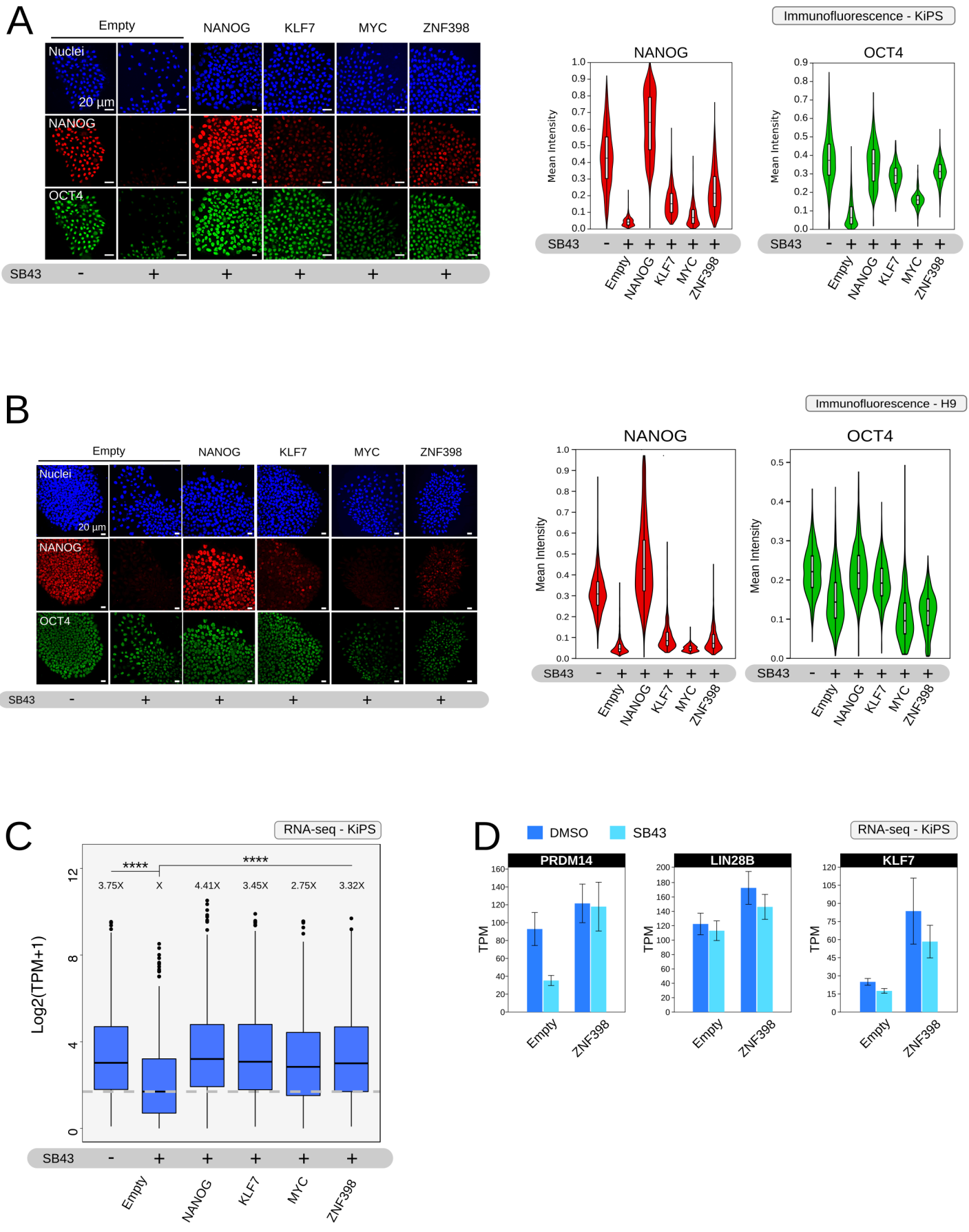

A
Empty
NANOG
KLF7
MYC
ZNF398
Nuclei
20 µm
NANOG
OCT4
SB43
Immunofluorescence - KiPS
NANOG
OCT4
Mean Intensity
B
Immunofluorescence - H9
C
RNA-seq - KiPS
****
3.75X
X
4.41X
3.45X
2.75X
3.32X
Log2(TPM+1)
D
DMSO
SB43
RNA-seq - KiPS
PRDM14
LIN28B
KLF7
TPM

Figure 7. ZNF398 represses hPSCs differentiation and epithelial to mesenchymal transition

(A) Transcriptome analysis of KiPS stably expressing an empty vector and treated with SB43 for 5 days. DOWN-regulated (Log2 fold-change < -1 and p-value < 0.01) and UP-regulated (Log2 fold-change > 1 and p-value < 0.01) genes are indicated in blue and orange respectively. Known TGF-beta targets (LEFTY1, LEFTY 2) serve as controls.

(B) Gene Ontology (GO) term analysis for biological processes of DOWN-regulated (left, blue bars) and UP-regulated (right, orange bars) revealed a statistically significant enrichment (p-value < 0.05) for genes involved in cell adhesion, epithelial to mesenchymal transition, organisation of extracellular matrix (highlighted in red) and neural development (highlighted in green).

(C) Heatmap for markers of neuroectodermal and epithelial character. RNA-seq data derived from KiPS stably expressing an empty vector (Empty) or ZNF398 and treated with DMSO or SB43 for 5 days. Z-Scores of row-scaled expression values (TPM) are shown. Orange and blue indicate high and low expression, respectively. Markers of epithelial (EPCAM, ESRP1, EZR) and neuroectodermal identity (SIX3, ENC1, PAX3) are highlighted.

(D) Validation of the siRNAs used in this work. H9 were transfected with the indicated siRNA and analysed after 3 days by qPCR for the indicated genes (NANOG or ZNF398). The siGFP was used as a further control. For each gene, 5 independent siRNAs have been tested and the 2 showing the best knockdown efficiency have been used as pool (grey bars).

(E) Gene expression analysis by qPCR of H9 transfected with the indicated siRNAs (a non-targeting siRNA (siCONTROL) or a pool of two validated siRNAs (siNANOG and siZNF398) and treated with or without SB43 for 5 days. See also Methods for details. Bars indicate mean and s.e.m. of 2 independent experiments. Expression was normalised to the mean of siCONTROL treated with SB43 samples.

FIGURE 7

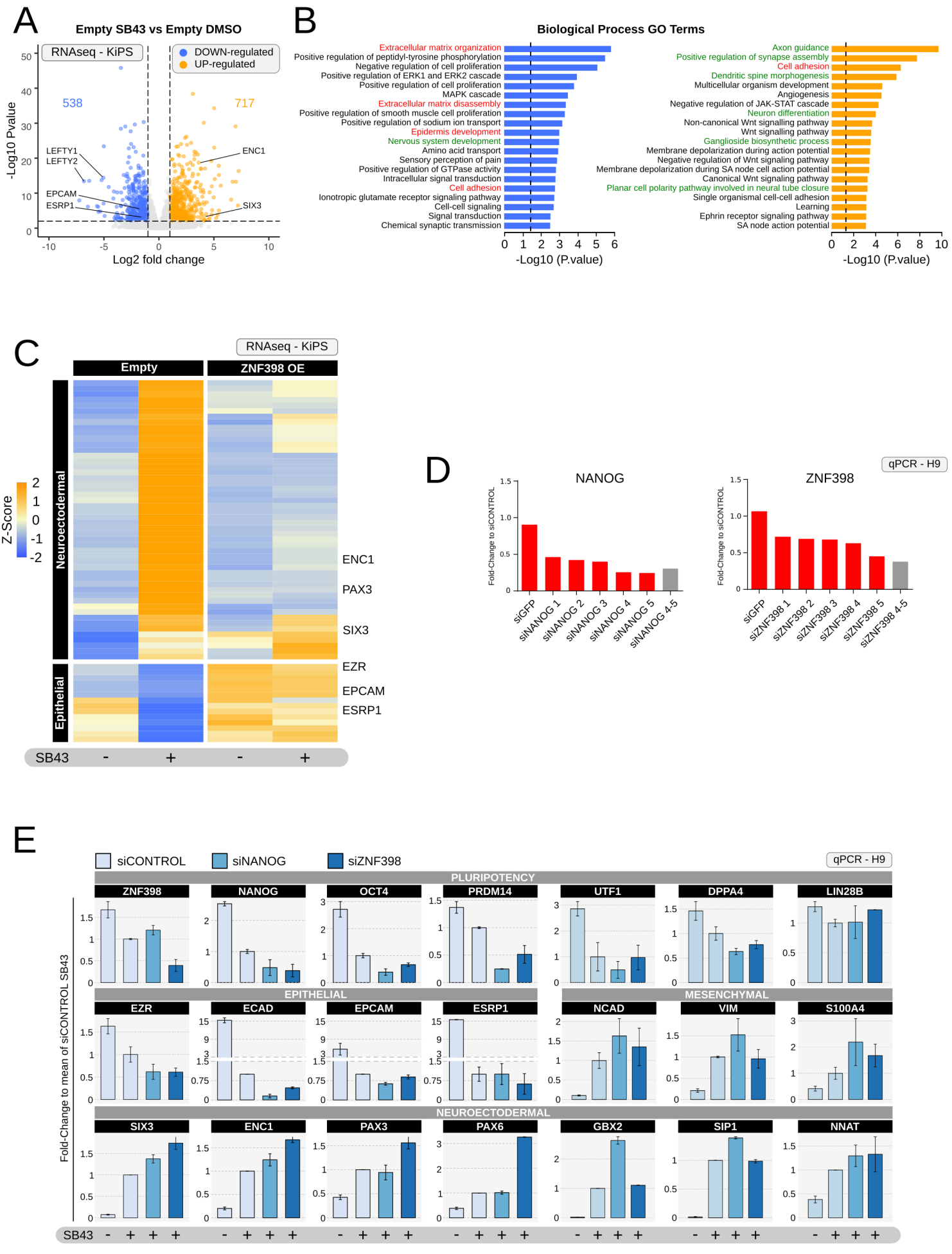

A
Empty SB43 vs Empty DMSO
RNAseq - KiPS
DOWN-regulated
UP-regulated
538
717
LEFTY1
LEFTY2
EPCAM
ESRP1
ENC1
SIX3
-Log10 Pvalue
Log2 fold change
B
Biological Process GO Terms
Extracellular matrix organization
Positive regulation of peptidyl-tyrosine phosphorylation
Negative regulation of cell proliferation
Positive regulation of ERK1 and ERK2 cascade
Positive regulation of cell proliferation
MAPK cascade
Extracellular matrix disassembly
Positive regulation of smooth muscle cell proliferation
Positive regulation of sodium ion transport
Epidermis development
Nervous system development
Amino acid transport
Sensory perception of pain
Positive regulation of GTPase activity
Intracellular signal transduction
Cell adhesion
Ionotropic glutamate receptor signaling pathway
Cell-cell signaling
Signal transduction
Chemical synaptic transmission
-Log10 (P.value)
Axon guidance
Positive regulation of synapse assembly
Cell adhesion
Dendritic spine morphogenesis
Multicellular organism development
Angiogenesis
Negative regulation of JAK-STAT cascade
Neuron differentiation
Non-canonical Wnt signalling pathway
Wnt signalling pathway
Ganglioside biosynthetic process
Membrane depolarization during action potential
Negative regulation of Wnt signaling pathway
Membrane depolarization during SA node cell action potential
Canonical Wnt signaling pathway
Planar cell polarity pathway involved in neural tube closure
Single organismal cell-cell adhesion
Learning
Ephrin receptor signaling pathway
SA node action potential
-Log10 (P.value)
C
RNAseq - KiPS
Empty
ZNF398 OE
Z-Score
Neuroectodermal
Epithelial
ENC1
PAX3
SIX3
EZR
EPCAM
ESRP1
SB43
D
qPCR - H9
NANOG
ZNF398
Fold-Change to siCONTROL
siGFP
siNANOG 1
siNANOG 2
siNANOG 3
siNANOG 4
siNANOG 5
siNANOG 4-5
siZNF398 1
siZNF398 2
siZNF398 3
siZNF398 4
siZNF398 5
siZNF398 4-5
E
siCONTROL
siNANOG
siZNF398
qPCR - H9
PLURIPOTENCY
ZNF398
NANOG
OCT4
PRDM14
UTF1
DPPA4
LIN28B
EPITHELIAL
EZR
ECAD
EPCAM
ESRP1
MESENCHYMAL
NCAD
VIM
S100A4
NEUROECTODERMAL
SIX3
ENC1
PAX3
PAX6
GBX2
SIP1
NNAT
Fold-Change to mean of siCONTROL SB43
SB43

Figure 8. The forced expression of ZNF398 moderates the differentiation process in EBs

(A) Experimental approach used to perform Embryoid Bodies (EBs) differentiation of KiPS stably expressing the empty vector (Empty) or ZNF398. Single cells were seeded on low attachment plate with E8 medium + ROCKi (10 uM). After 2 days, the medium was changed with DMEM/20% FBS, a medium that induces differentiation. Medium was changed every 2 days. EBs of empty vector (Empty) or ZNF398 expressing cells were collected at 3 different time points (Day 0, Day 5 and Day 15) of differentiation and analysed by qPCR. See also Methods for details.

(B) Gene expression analysis by qPCR of EBs differentiation of KiPS stably expressing an empty vector (Empty, light orange) or ZNF398 (dark orange) analysed at 3 different time points (Day 0, Day 5 and Day 15) of differentiation. Bars indicate mean and s.e.m of 5 independent experiments, shown as dots. Expression was normalised to the mean of Empty Day 0 sample. Unpaired two-tailed Mann–Whitney *U* test.

FIGURE 8

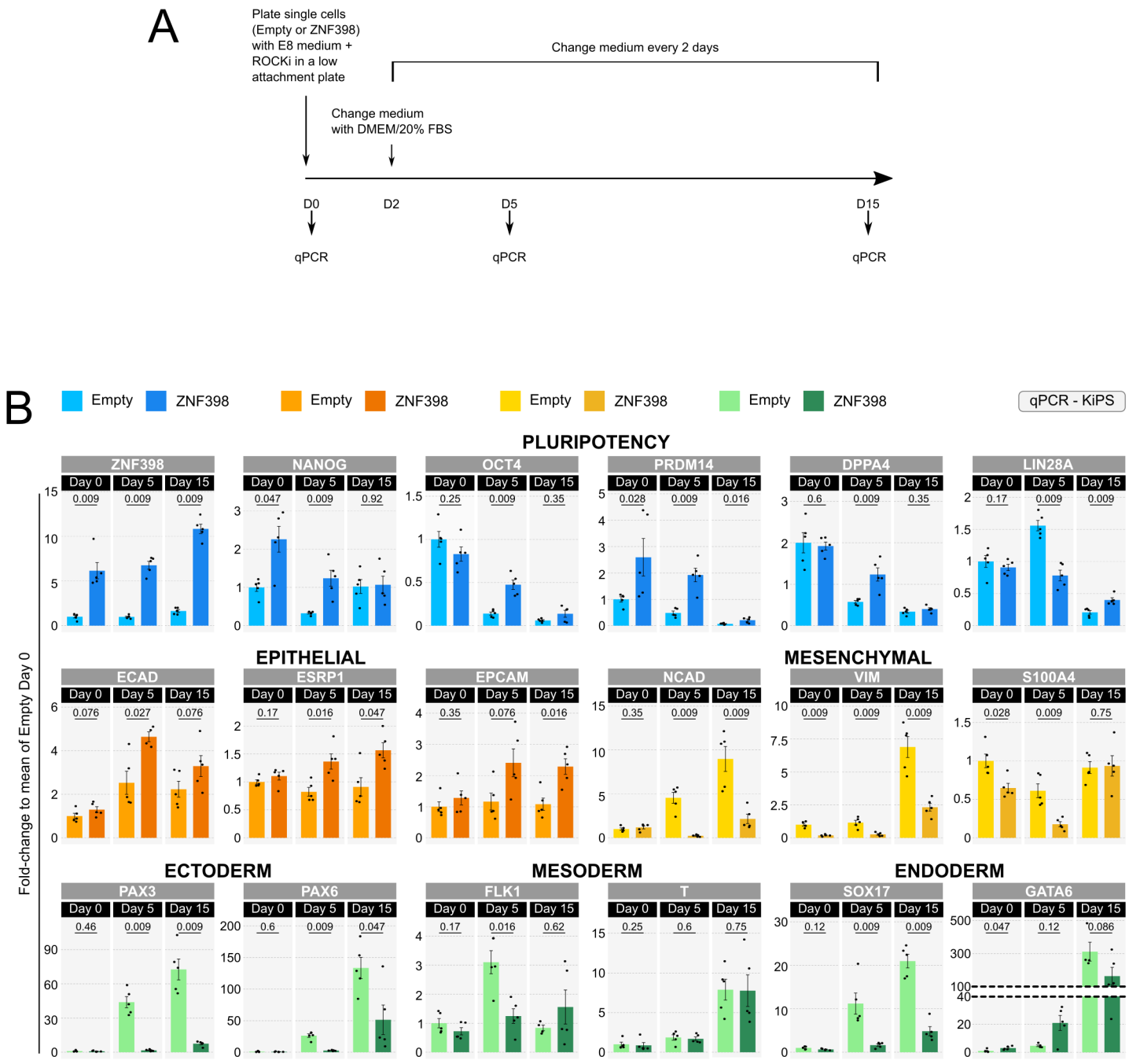

A
Plate single cells (Empty or ZNF398) with E8 medium + ROCKi in a low attachment plate
Change medium every 2 days
Change medium with DMEM/20% FBS
D0
D2
D5
D15
qPCR
qPCR
qPCR
B
Empty
ZNF398
Empty
ZNF398
Empty
ZNF398
Empty
ZNF398
qPCR - KiPS
PLURIPOTENCY
ZNF398
NANOG
OCT4
PRDM14
DPPA4
LIN28A
EPITHELIAL
MESENCHYMAL
ECAD
ESRP1
EPCAM
NCAD
VIM
S100A4
ECTODERM
MESODERM
ENDODERM
PAX3
PAX6
FLK1
T
SOX17
GATA6
Day 0
Day 5
Day 15
Fold-change to mean of Empty Day 0

Figure 9. ZNF398 acts as a transcriptional activator in concert with SMAD3 and EP300

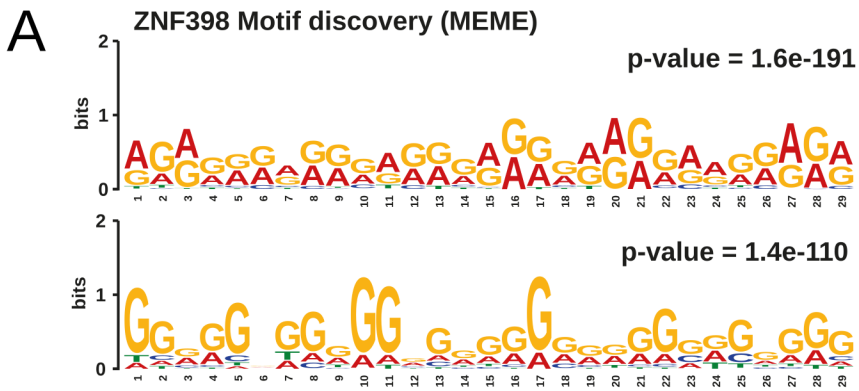
(A) DNA binding motifs associated with ZNF398 peaks.

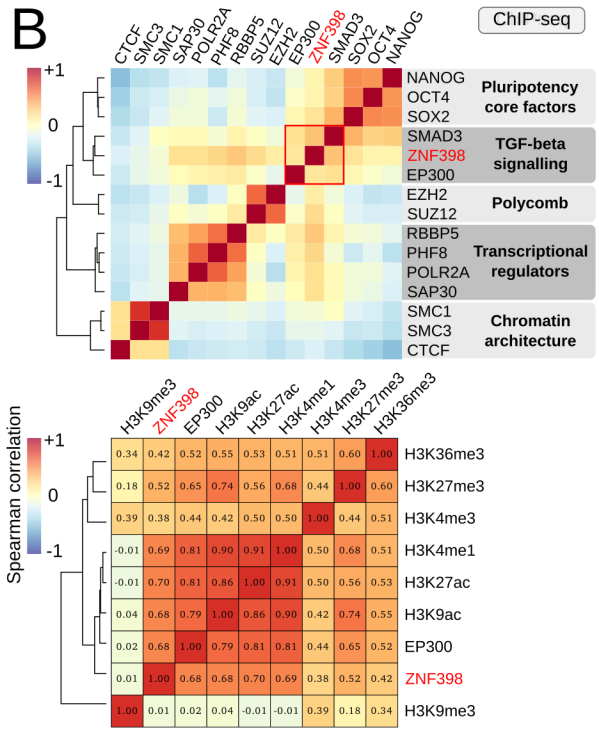
(B) Top: Hierarchical clustering of 15 genome-wide binding profiles (available genome-wide binding profiles from CODEX, Sánchez-Castillo et al., 2015). Normalised Pointwise Mutual Information (NPMI) between each pair of samples were used to display all pairwise binding overlaps in a clustered heatmap. Colours in the heatmap show the level of overlap for each pair of samples (red, all binding sites overlapped; yellow, overlap expected by chance; blue, mutually exclusive binding). Bottom: Hierarchical clustering of pairwise Spearman correlation of ZNF398 and ChiP-seq datasets indicated. Colours indicate the level of correlation (red indicates perfect correlation, blue indicates anticorrelation). ZNF398 clusters together with active histone marks.

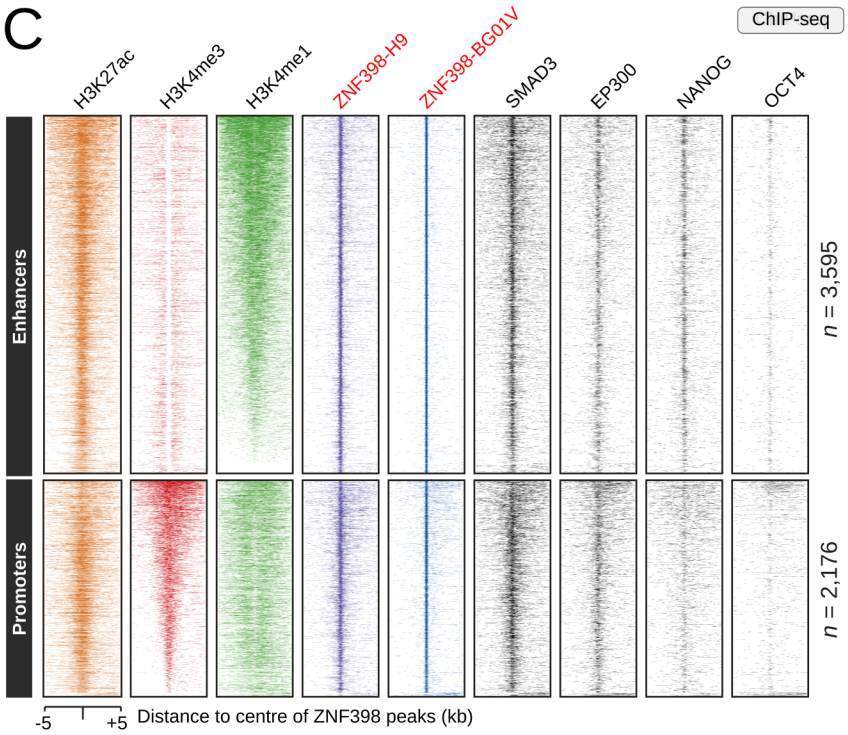
(C) Binding plots show the location of ZNF398 (obtained in two hPSC lines, H9 and BG01V) and H3K27ac, H3K4me3, H3K4me1, SMAD3, EP300, NANOG and OCT4. 5,771 sites are displayed within a 10kb window centred around ZNF398 peaks. Note the presence of ZNF398 both at active enhancers (H3K4me1 positive) and active promoters (H3K4me3 positive).

(D) ZNF398 binding and absolute expression (TPM) of a pluripotency (LIN28B, left), an epithelial (ESRP1, centre) markers and a canonical SMAD3 target (LEFTY1, right). Gene tracks: binding of ZNF398 in 2 different hPSC lines are shown. EP300, SMAD3, H3K27ac tracks show colocalization with ZNF398 binding sites. Barcharts: gene expression analysis by RNA-seq of KiPS stably expressing an empty vector (Empty) or ZNF398. Bars indicate the mean and the s.e.m of at least 2 independent experiments. Absolute expression is reported as TPM.

FIGURE 9


A
ZNF398 Motif discovery (MEME)
p-value = 1.6e-191
p-value = 1.4e-110
bits


B
ChIP-seq
CTCF SMC3 SMC1 SAP30 POLR2A PHF8 RBBP5 SUZ12 EZH2 EP300 ZNF398 SMAD3 SOX2 OCT4 NANOG
NANOG
OCT4
SOX2
Pluripotency core factors
SMAD3
ZNF398
EP300
TGF-beta signalling
EZH2
SUZ12
Polycomb
RBBP5
PHF8
POLR2A
SAP30
Transcriptional regulators
SMC1
SMC3
CTCF
Chromatin architecture
Spearman correlation
H3K9me3 ZNF398 EP300 H3K9ac H3K27ac H3K4me1 H3K4me3 H3K27me3 H3K36me3
H3K36me3
H3K27me3
H3K4me3
H3K4me1
H3K27ac
H3K9ac
EP300
ZNF398
H3K9me3


C
ChIP-seq
H3K27ac H3K4me3 H3K4me1 ZNF398-H9 ZNF398-BG01V SMAD3 EP300 NANOG OCT4
Enhancers
Promoters
n = 3,595
n = 2,176
-5 +5 Distance to centre of ZNF398 peaks (kb)

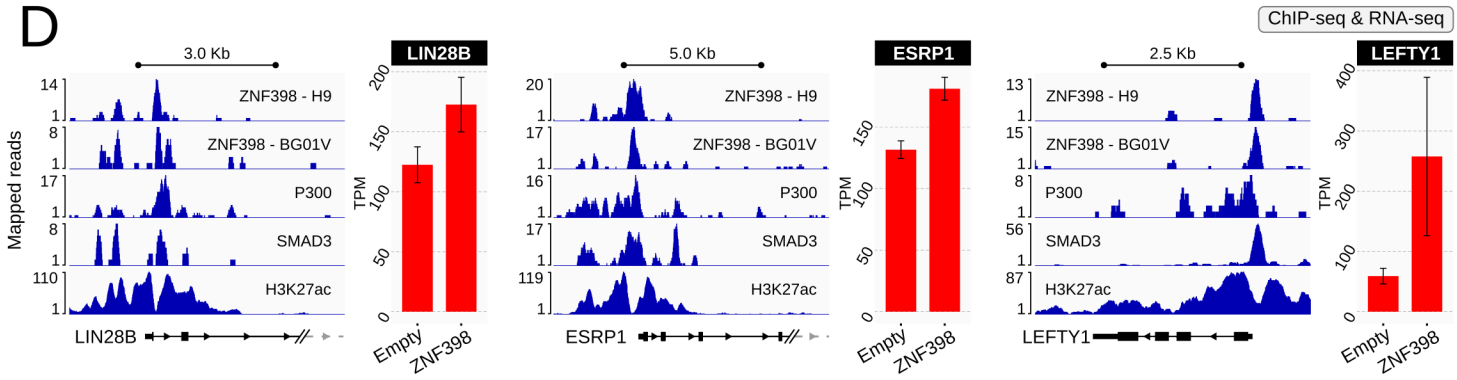

D
ChIP-seq & RNA-seq
3.0 Kb
5.0 Kb
2.5 Kb
ZNF398 - H9
ZNF398 - BG01V
P300
SMAD3
H3K27ac
Mapped reads
LIN28B
ESRP1
LEFTY1
TPM
Empty
ZNF398

Figure 10. Effect of ZNF398 on TGF-beta signaling

(A) Scatter plot showing RNA-seq data from KiPS stably expressing an empty vector (Empty) or ZNF398. DOWN-regulated (Log2FC < -0.5) and UP-regulated (Log2FC > 0.5) genes are indicated in blue and orange respectively. Known SMAD3 targets that are also TGF-beta signalling pathway components are highlighted.

(B) Box plot showing mean-normalised expression levels of 81 SMAD3 targets UP-regulated by TGF-beta induction identified in Figure 3A. 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 3 samples. Unpaired two-tailed *t*-test.

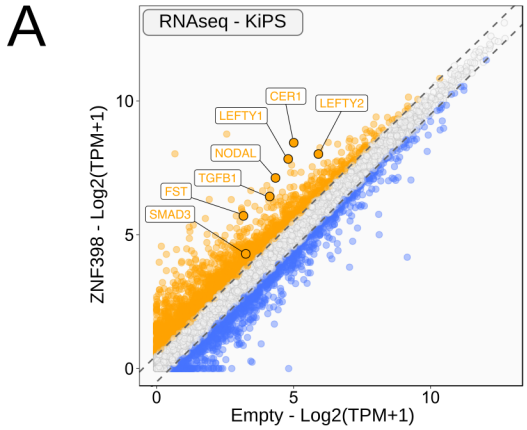
(C) LEFTY1 gene expression analysis by qPCR of HES2 (light orange bars) and KiPS (dark orange bars) stably expressing an empty vector (Empty) or ZNF398 and treated with or without SB43 for 5 days. Bars indicate mean and s.e.m of 4 independent experiments. Expression was normalised to the Empty SB43 samples and shown on a logarithmic scale. Unpaired two-tailed Mann–Whitney *U* test.

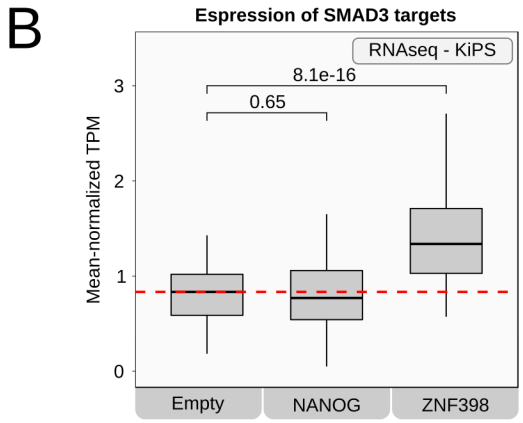
(D) Top: Representative images of immunostaining for total SMAD2/3 of KiPS stably expressing an empty vector (Empty) control or ZNF398 treated with SB43 for 16 hours, or after TGF-beta induction (2ng/ml) for 15, 30 and 60 minutes (min). Nuclei were identified by DAPI staining. Representative images of 2 independent experiments are shown. See also Figure 10E.
Bottom: line charts showing SMAD2/3 intensity in arbitrary units across hPSCs (dashed yellow line). On X axis, distance is expressed in pixels where 0 is the estimated nuclear centre and +/-20 correspond to the cytoplasm. For each condition, 48 cells from 6 randomly selected fields were analysed to calculate SMAD2/3 median intensity and 1st-3rd quartile interval, respectively represented by the black line and the shaded area. “U”-shaped trend indicates that SMAD2/3 is mainly cytoplasmic

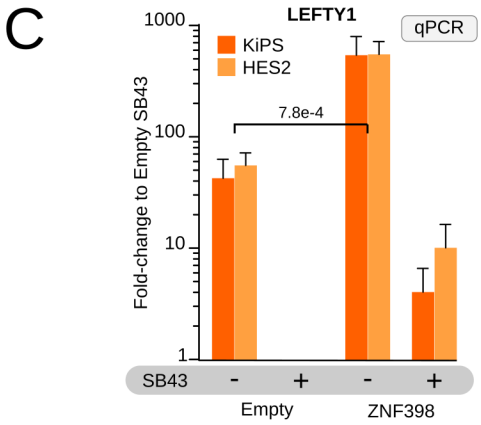
(Empty-SB43), conversely when SMAD2/3 is mostly nuclear the trend is flipped (ZNF398-60min). Flat trend suggests that SMAD2/3 is equally distributed between cytoplasm and nucleus (ZNF398-15min).

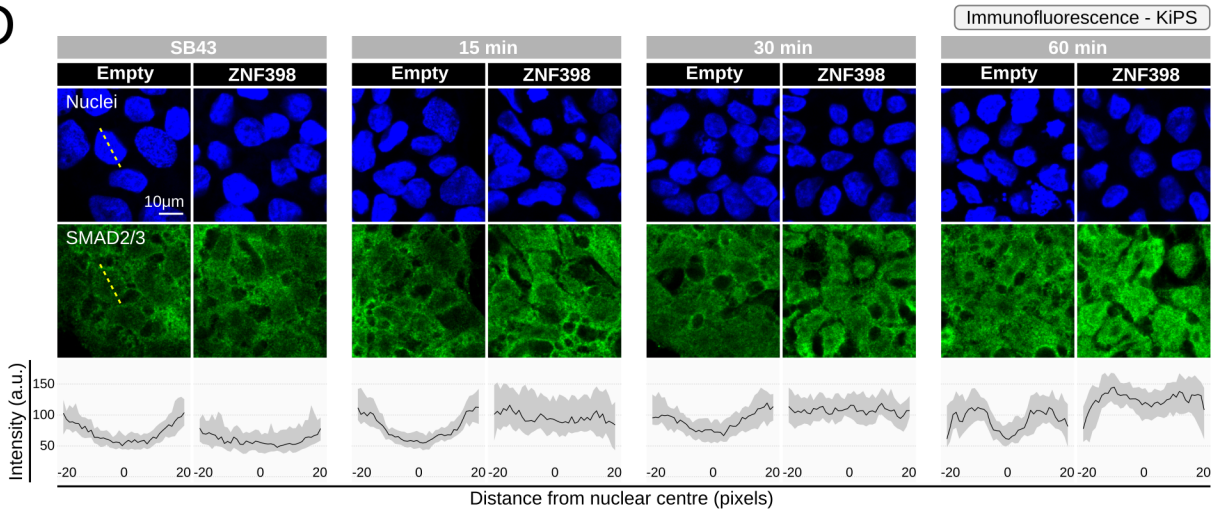
(E) Representative images of immunostaining for total SMAD2/3 of KiPS stably expressing the empty vector (Empty) or ZNF398 in presence of SB43, TGF-beta induction (2ng/ml) for 15, 30 and 60 minutes (min). Nuclei were identified by DAPI staining. Representative images of 2 independent experiments are shown. Dashed yellow box highlight the close up shown in Figure 10D.

FIGURE 10


A
RNAseq - KiPS
CER1
LEFTY2
LEFTY1
NODAL
TGFB1
FST
SMAD3
ZNF398 - Log2(TPM+1)
Empty - Log2(TPM+1)
0
5
10


B
Espression of SMAD3 targets
RNAseq - KiPS
8.1e-16
0.65
Mean-normalized TPM
Empty
NANOG
ZNF398


C
LEFTY1
qPCR
KiPS
HES2
7.8e-4
Fold-change to Empty SB43
SB43
Empty
ZNF398


D
Immunofluorescence - KiPS
SB43
15 min
30 min
60 min
Empty
ZNF398
Nuclei
10µm
SMAD2/3
Intensity (a.u.)
Distance from nuclear centre (pixels)

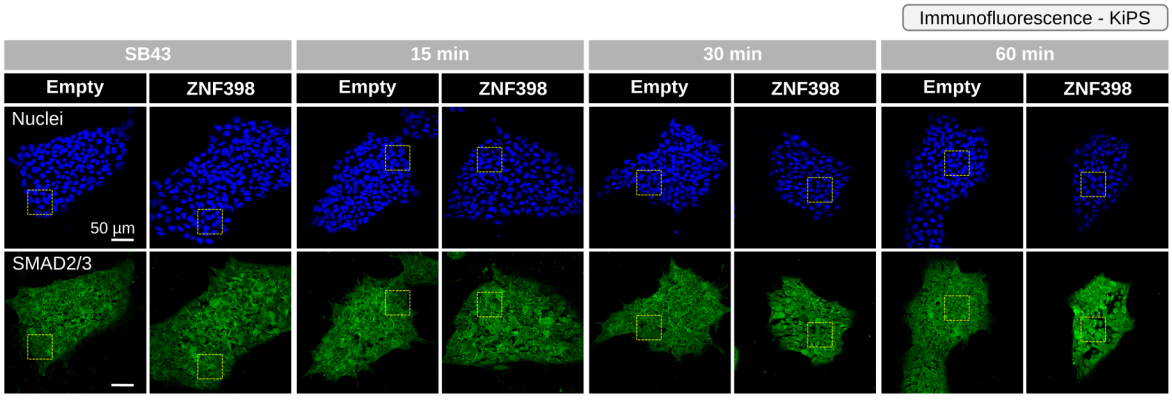

E
Immunofluorescence - KiPS
SB43
15 min
30 min
60 min
Empty
ZNF398
Nuclei
50 µm
SMAD2/3

Figure 11. ZNF398 is required for somatic cell reprogramming

(A) Experimental strategy for reprogramming by delivery of OSKMNL (OCT4, SOX2, MYC, NANOG, LIN28A) mRNAs in combination with siRNAs in order to test the requirement of ZNF398 for reprogramming. Human BJ fibroblasts were daily transfected with OSKMNL mRNAs between day 1 and 8 and also with siRNAs (siCONTROL or siZNF398) at day 1, 3, 5. From day 8, fresh mTeSR medium was added daily to help stabilisation of hiPSCs colonies. At day 6 and day 14, colonies were analysed as indicated. See also Methods for details.

(B) Optimisation of transfection conditions for mRNAs in combination with siRNAs. Human BJ foreskin fibroblasts were transfected with a mRNA encoding for GFP in combination with siRNAs. Representative images of the indicated conditions after 3 days of transfection are shown. Fibroblasts that were not transfected served as a control. Note that the majority of fibroblasts express GFP protein in presence of siCONTROL, while GFP signal is strongly reduced by siGFP.

(C) Cell morphology during reprogramming. At day 6 of reprogramming, fibroblasts normally convert from a spindle-like to an epithelial-like morphology (see the no siRNA and CONTROL conditions), but the knockdown of ZNF398 reduced such conversion. At day 14 of reprogramming hiPSCs colonies were obtained in all the three conditions, but were reduced in number and size upon knockdown of ZNF398.

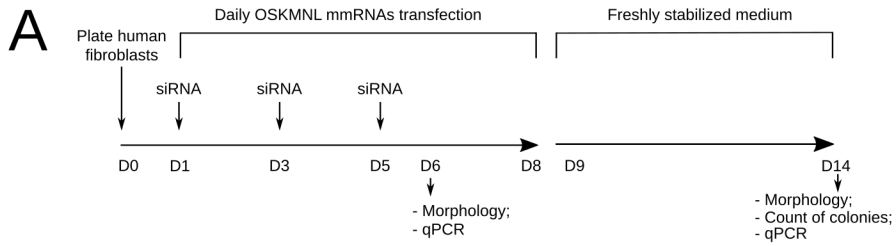
(D) Number of colonies per channel obtained at day 14 in the indicated conditions. Bars indicate the mean and s.e.m. of 4 independent experiments, shown as dots. Unpaired two-tailed Mann–Whitney *U* test.

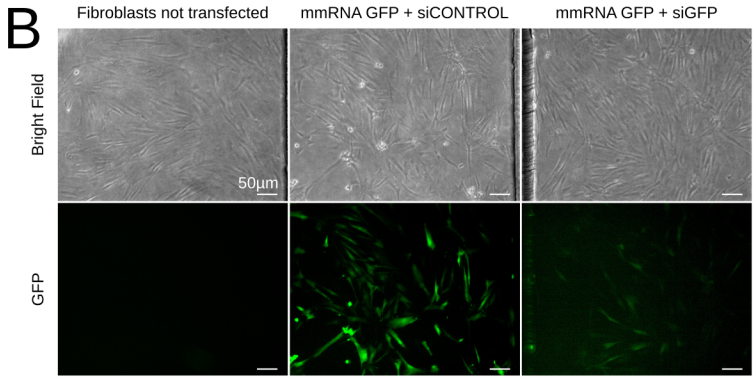
(E) Gene expression analysis by qPCR of hiPSCs (KiPS) (light grey) and fibroblasts (dark grey) serving as controls, and fibroblasts transfected with OSKMNL mRNAs and siCONTROL (light orange) or siZNF398 (dark orange) at day 6 and day 14. ZNF398

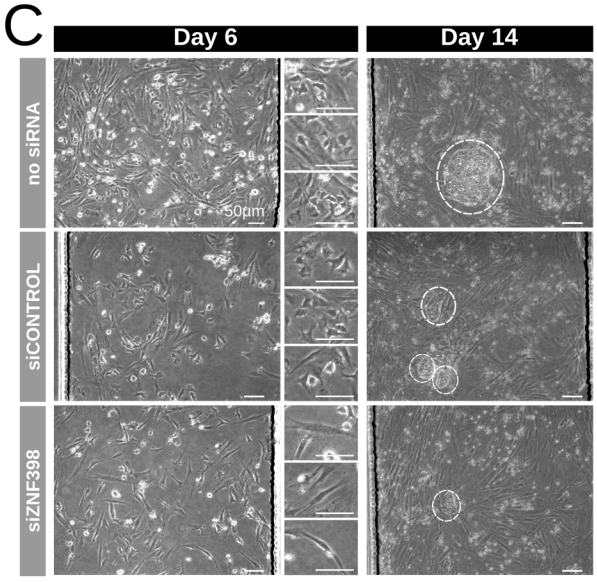
levels are transiently reduced at day 6. Bars indicate the mean of 2 independent experiments. Expression was normalised to the mean of KiPS samples.

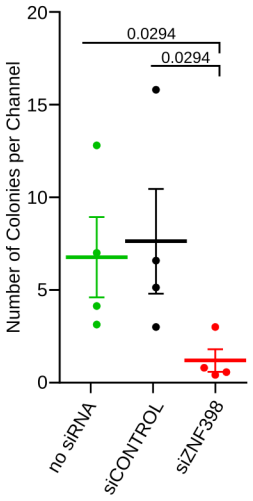
(F) Heatmap showing gene expression analysis by qPCR for markers of pluripotency and epithelial character during reprogramming experiments. Z-Scores of expression values from 2 independent experiments. Red and blue indicate high and low expression, respectively.

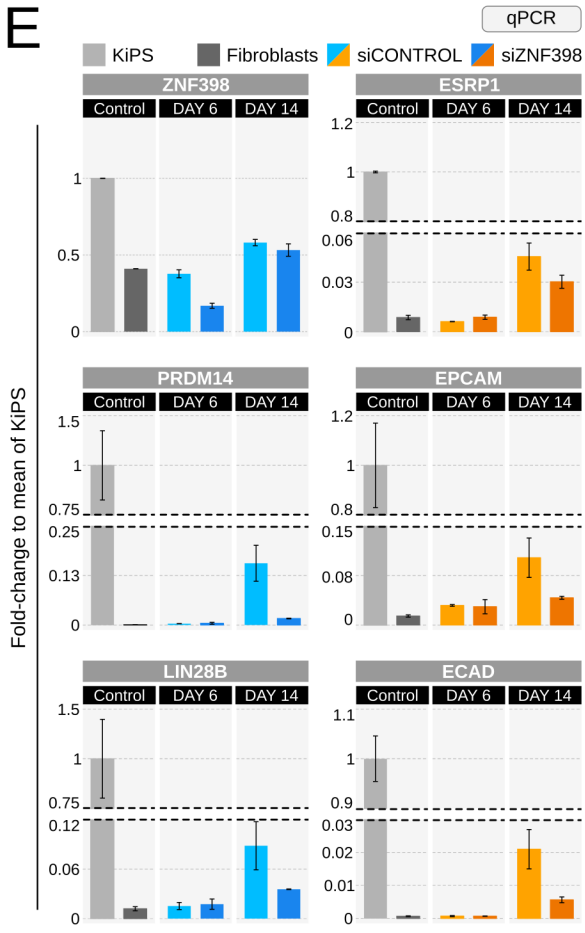
FIGURE 11


A
Daily OSKMNL mmRNAs transfection
Freshly stabilized medium
Plate human fibroblasts
siRNA
siRNA
siRNA
D0
D1
D3
D5
D6
D8
D9
D14
- Morphology;
- qPCR
- Morphology;
- Count of colonies;
- qPCR


B
Fibroblasts not transfected
mmRNA GFP + siCONTROL
mmRNA GFP + siGFP
Bright Field
GFP
50µm


C
Day 6
Day 14
no siRNA
siCONTROL
siZNF398
50µm


D
Number of Colonies per Channel
0.0294
0.0294
no siRNA
siCONTROL
siZNF398


E
qPCR
KiPS
Fibroblasts
siCONTROL
siZNF398
Fold-change to mean of KiPS
ZNF398
ESRP1
PRDM14
EPCAM
LIN28B
ECAD
Control
DAY 6
DAY 14

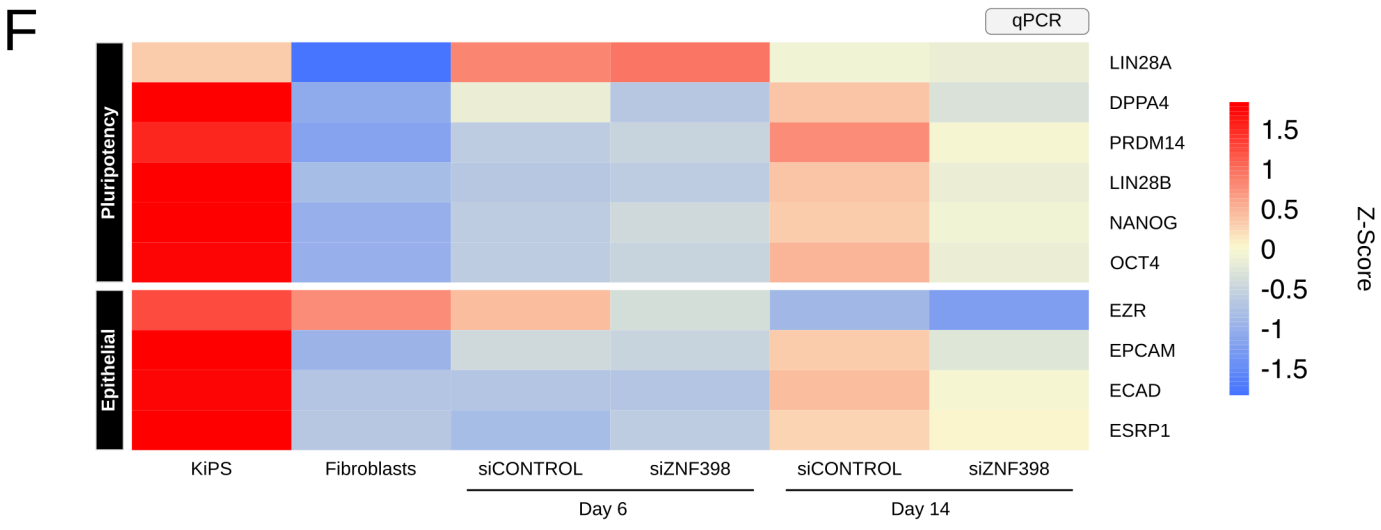

F
qPCR
Pluripotency
Epithelial
LIN28A
DPPA4
PRDM14
LIN28B
NANOG
OCT4
EZR
EPCAM
ECAD
ESRP1
KiPS
Fibroblasts
siCONTROL
siZNF398
siCONTROL
siZNF398
Day 6
Day 14
Z-Score

Figure 12. ZNF398 promotes 3D self-organization

(A) Experimental approach used to test the self-organization capacity of KiPS stably expressing the empty vector (Empty) or ZNF398 with or without SB43. Single cells were seeded on Matrigel with E8 medium + 5% of Matrigel (v/v) with or without SB43. After 4 days, 3D structures were analysed as indicated. See also Methods for details.

(B) Representative images of immunostaining for NANOG and Phalloidin markers or DAPI and beta-catenin of KiPS (wild-type) after 4 days of self-organization. The 3D structures were characterized by a columnar epithelium (Phalloidin and BETA-CATENIN) with basally located nuclei, expressing the pluripotency marker NANOG.

(C) Representative morphology of KiPS stably expressing the empty vector (Empty) or ZNF398 in presence of DMSO or SB43. SB43 abolishes the formation of epithelial organization, which are rescued by forced expression of ZNF398.

(D) Bar plot showing the number of structures obtained after 4 days of self-organization. Light blue bars indicate epithelial organized structures, while the orange bars indicate disorganised structures. Bars indicate mean and s.e.m. of 3 independent experiments, shown as dots.

(E) Gene expression analysis by qPCR of KiPS stably expressing the empty vector control (Empty) or ZNF398 with or without SB43 after 4 days of self-organization. Bars indicate mean and s.e.m. of 4 independent experiments. Expression was normalised to the mean of Empty DMSO samples. Unpaired two-tailed Mann–Whitney *U* test.

(F) Heatmap showing gene expression analysis by qPCR for markers of pluripotency, epithelial and mesenchymal identity of self-organisation experiments. Human PSCs, either expressing ZNF398 or an empty vector, were analysed when cultured under

conventional conditions (2D) and after 4 days of self-organisation (3D) in presence of SB43 or DMSO. Values are expressed as fold-change relative to the mean of Empty 2D samples from 4 independent experiments. Red and blue indicate high and low expression, respectively.

FIGURE 12

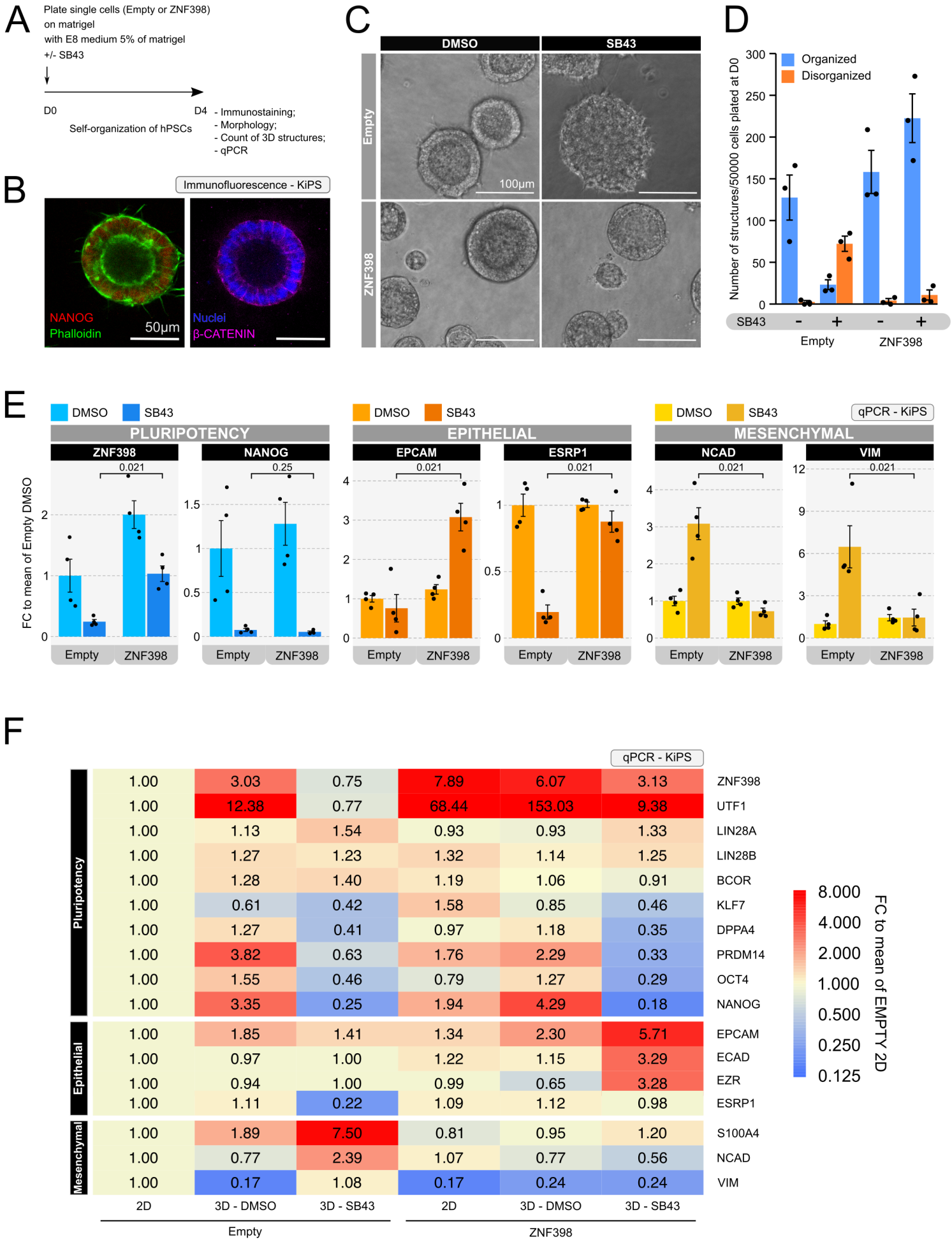
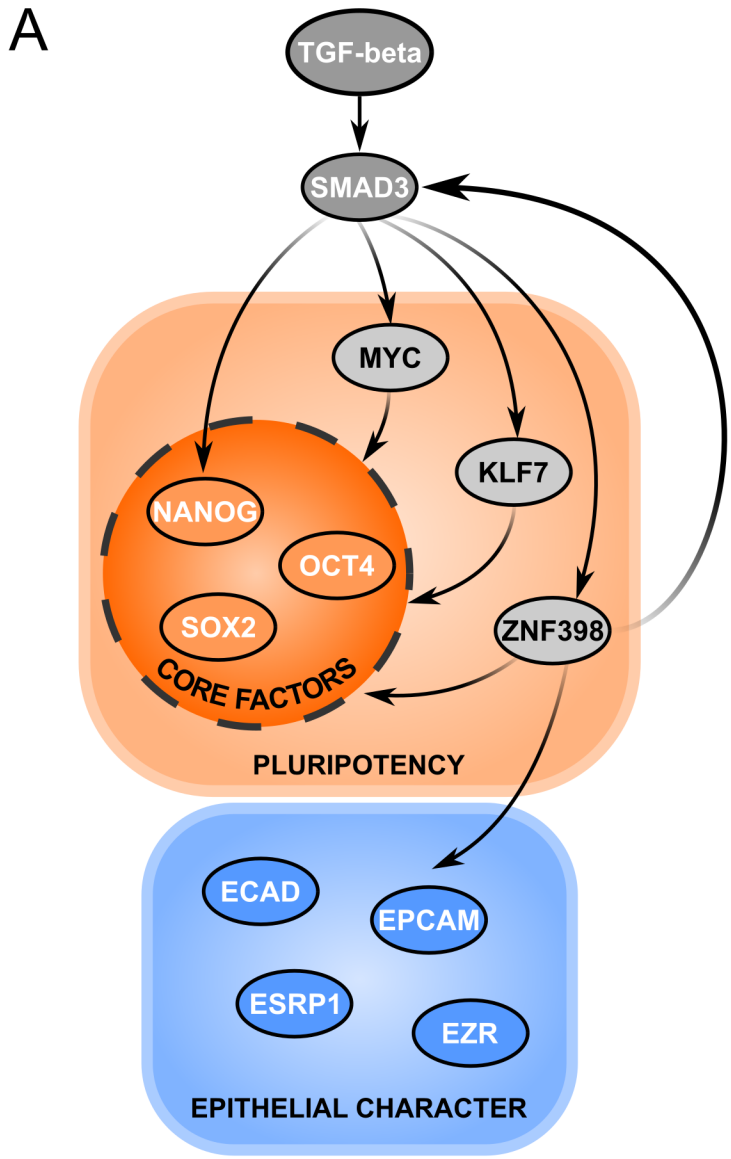

A
Plate single cells (Empty or ZNF398)
on matrigel
with E8 medium 5% of matrigel
+/- SB43
D0
D4
Self-organization of hPSCs
- Immunostaining;
- Morphology;
- Count of 3D structures;
- qPCR
B
Immunofluorescence - KiPS
NANOG
Phalloidin
50µm
Nuclei
β-CATENIN
C
DMSO
SB43
Empty
ZNF398
100µm
D
Number of structures/50000 cells plated at D0
Organized
Disorganized
SB43
Empty
ZNF398
E
DMSO
SB43
qPCR - KiPS
PLURIPOTENCY
EPITHELIAL
MESENCHYMAL
ZNF398
NANOG
EPCAM
ESRP1
NCAD
VIM
0.021
0.25
FC to mean of Empty DMSO
Empty
ZNF398
F
qPCR - KiPS
Pluripotency
Epithelial
Mesenchymal
ZNF398
UTF1
LIN28A
LIN28B
BCOR
KLF7
DPPA4
PRDM14
OCT4
NANOG
EPCAM
ECAD
EZR
ESRP1
S100A4
NCAD
VIM
2D
3D - DMSO
3D - SB43
Empty
ZNF398
FC to mean of EMPTY 2D

Figure 13. Graphical summary

(A) Diagram representing the transcription factors induced by TGF-beta, among which ZNF398 is crucial for the maintenance of pluripotency and the epithelial character of hPSCs.

FIGURE 13


A
TGF-beta
SMAD3
MYC
KLF7
NANOG
OCT4
SOX2
CORE FACTORS
ZNF398
PLURIPOTENCY
ECAD
EPCAM
ESRP1
EZR
EPITHELIAL CHARACTER

METHODS

Cell culture

hESCs (HES2, H9 and BG01V) and hiPSCs (KiPS) were cultured in feeder-free on pre-coated plates with 0.5% growth factor-reduced Matrigel (CORNING Cat. 356231) (vol/vol in PBS with MgCl2/CaCl2_Sigma-Aldrich D8662) in E8 medium (made in-house according to Chen et al., 2011) or in mTeSR (StemCell Technologies 05850) at 37°C, 5% CO_2, 5% O_2. Cells were passaged every 3-4 days at a split ratio of 1:8 following dissociation with EDTA solution (0.5 mM in PBS without MgCl2/CaCl2_Sigma-Aldrich D8662, pH 8.0, ThermoFisher). The human foreskin fibroblasts BJ (passage 12, ATCC, CRL-2522) were cultured in DMEM/F12 (Sigma-Aldrich D6421) with 10% fetal bovine serum (FBS; Sigma-Aldrich F7524) at 37°C, 5% CO_2, 21% O_2.

H9 line (WA09) and HES2 were obtained from and used under authorisation from WiCell Research Institute. KiPS were kindly provided by Austin Smith's laboratory (Takashima et al., 2014). BG01V were obtained from Fisherscentific (Cat. 10713674). 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 Figure 3D 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 SB431542 (SB43, Axon Medchem 1661) overnight. The morning after, TGF-beta signaling was re-induced by changing medium with mTeSR1 for 1 hour or for 4 hours. For the validation experiments on feeders, KiPS were plated on MEF and the next day, cells were treated with DMSO (Sigma-Aldrich D2650) or with 10 µM SB43 overnight. The morning after, cells were treated with 2ng/ml of TGFB1 (PEPROTECH Cat.100-21) or with 25 ng/ml of Activin A (QKINE Cat. Qk001) for 1 hour.

Generation of hPSCs stably expressing genes of interest

Stable transgenic hPSCs expressing candidates were generated by transfecting cells with PiggyBac (PB) transposon plasmids with piggyBac transposase expression vector pBase. The PiggyBac plasmid is able to stably integrate the gene of interest in the genome. PB retrotransposon is a mobile genetic element that is efficiently transposed from plasmid vectors to chromosomal DNA through a "copy and paste" mechanism. For the transposition is required the activity of an enzyme transposase pBase that is able to recognize specific inverted terminal repeat sequences (ITRs) located at both ends of the transposon vector and it moves the insert from the original position into a genomic site TTAA. This system has the following advantages: it allows the integration of the gene without the use of viral vectors, there is no limit as regards the size of the insert and finally the system is reversible; indeed, it is possible to re-transfect the cells with the transposase allowing the removal of the transposon stably integrated in the genome.

In order to generate the PB plasmids, the candidates (NANOG, ZNF398, KLF7, MYC, ETS2, OTX2, ID1, BCOR and PRDM14) were amplified from cDNA and cloned into a pENTR2B donor vector. Then, the transgenes were gateway cloned into the same destination vector containing PB-CAG-DEST-bghpA and pGK-Hygro selection cassette.

For DNA transfection, 250,000 hPSCs were dissociated as single cells with TrypLE (ThermoFisher) and were cotransfected with PB constructs (550 ng) and pBase plasmid (550 ng) using FuGENE HD Transfection (Promega, Cat. E2311), following the protocol for reverse transfection. For one well of a 12-well plate, we used 3,9 µl of transfection reagent, 1 µg of plasmid DNA, and 250,000 cells in 1 ml of E8 medium with 10µM Y27632 (ROCKi, Rho-associated kinase (ROCK) inhibitor_Axon 1683). The medium was changed after overnight incubation and Hygromycin B (200 µg/ml; Invitrogen Cat. 10687010) was added after 48 hours.

For the overexpression experiments, hPSCs stably expressing an empty vector or the candidates were plated. The next day, cells were treated with DMSO or 10µM SB43 for 5 days and then analysed as indicated in Figure 4A.

siRNA transfection

For siRNA transfection, hPSCs were plated on Matrigel-coated 24 well-plate as clusters (2500-5000 clusters for one well of a 24-well plate) in E8 medium with 10µM ROCKi. After 4 hours, siRNAs were transfected at a final concentration of 20 nM using Stemfect^{TM} RNA Transfection Kit (STEMGENT Cat. 00-0069), following the protocol for forward transfection. For a 24-well plate (2 cm^2), I used 0.52 µl of transfection reagent, 2 µl of 10µM siRNA solution and 25 µl of transfection buffer. After waiting 20 minutes, I mixed the transfection mix with 1ml of E8 medium. The medium was changed after overnight incubation. See Table 1 for sequences of the siRNAs used.

Embryoid bodies (EB) differentiation assay

KiPS stably expressing an empty vector or ZNF398 were detached as clumps with EDTA and plated on ultra low attachment surface plates (CORNING 3473) in E8 medium with 10µM ROCKi. After 2 days, E8 medium was substituted with DMEM, 20% FBS, 200 mM L-glutamine, 1% NEAA and 0.1 mM 2-mercaptoethanol. Medium was changed every 2 days.

3D culture of hPSCs

hPSCs were suspended as single cells using TrypLE^{TM} Select (1X) (Thermo Fisher 12563-029). I covered a well of a Chamber slide (Thermo scientific 177402, 0.8 cm^2 each) with 40µl of pure ice-cold growth factor reduced Matrigel (CORNING Cat. 356231) and incubated for 2 min at 37°C to allow Matrigel to solidify. I resuspended 50,000 cells (for each well of a chamber slide) in 250 µl of E8 medium with 10µM ROCKi. 10 mins after plating, I removed the medium and replaced with 250 µl of E8 + 10µM ROCKi containing 5% Matrigel. After 24 hours, I changed the medium (E8

medium containing 5% of Matrigel v/v). The 3D structures were grown at 37°C, 5% CO_2, 21% O_2.

Reprogramming

All reprogramming experiments were performed in microfluidics as previously described (Gagliano et al., 2019) in hypoxia conditions (37°C, 5% CO_2, 5% O_2). The protocol for reprogramming experiments were optimized to transfect siRNA in order to test the requirement of ZNF398 for reprogramming. Briefly, microfluidic channels were coated with 25 µg ml^{-1} Vitronectin (ThermoFisher, A14700) for 1 hour at room temperature (RT) and fibroblasts were seeded at day 0 at 30 cells per mm^2 in DMEM/10% FBS. On day 1, 9 hours before the first mRNAs transfection, I applied E6 medium (made in-house according to (Chen et al., 2011)) including FGF2 100ng ml^{-1} (QKINE Cat. no Qk002 recombinant zebrafish FGF2), 20% KSR (Gibco, 10828028), ROCKi 5µM and LSD1i 0.1µM (RN-1, EMD Millipore Cat. no 489479). Cells were transfected daily at 6PM and fresh medium was given daily at 9AM. The transfection mix was prepared according to the StemMACS™ mRNA Transfection Kit (Miltenyi Biotec, 130-104-463) and Stemgent StemRNA-NM Reprogramming Kit (Reprocell, 00-0076) (OSKMNL NM-RNA (Not-Modified RNA) and EKB NM-RNA) and prepare the RNA mix according to the manufacturer's instructions. siRNAs were transfected at final concentration of 20 nM at day 1, day 3 and day 5 (see Table 1 for sequences of the siRNAs used) together with mRNAs. During the 8 days of transfection, the dose of mRNAs was gradually increased according to cell proliferation rate and transfection-induced cell mortality.

Immunofluorescence and stainings

Immunofluorescence analysis was performed on 1% Matrigel-coated glass coverslip in wells. Cells were fixed in 4% Formaldehyde (Sigma-Aldrich Cat. 78775-500ML) in PBS for 10 minutes at RT, washed in PBS, permeabilized for 1 hour in PBS + 0.3% Triton X-100 (PBST) at RT, and blocked in PBST + 5% of Horse serum (ThermoFisher Cat. 16050-122) for 5 hours at RT. Cells were incubated overnight at 4°C with primary antibodies (See Supplementary Table 2) in PBST+ 3% of Horse serum. After washing

with PBS, cells were incubated with secondary antibodies (Alexa, Life Technologies) (Supplementary Table 2) for 45 minutes at RT. Cells were mounted with Fluoroshield™ with DAPI (Sigma-Aldrich, cat. F6057-20ML).

For immunofluorescence on 3D structures (Figure 12 B), 3D structures were fixed in 4% Formaldehyde in PBS for 45 minutes at RT, washed in PBS, permeabilized for 1 hour in PBS + 0.3% Triton X-100 (PBST) at RT, and blocked in PBST + 5% of Horse serum for 5 hours at RT. 3D structures were incubated overnight at 4°C with primary antibodies (see Table 2) in PBST + 3% of Horse serum. After washing with PBS, 3D structures were incubated with secondary antibodies (Alexa, Life Technologies) (see Table 2) overnight at 4°C. Nuclei were stained with Hoechst 33342 (ThermoFisher Cat. 62249). Images were acquired with a Zeiss LSN700 or Leica SP5 confocal microscope equipped with a charge-coupled device camera using ZEN 2 or Leica LAS AF software.

For alkaline phosphatase staining, cells were fixed with a citrate–acetone–formaldehyde solution and stained using an alkaline phosphatase detection kit (Sigma-Aldrich, cat. 86R-1KT). Plates were scanned using an Epson scanner and scored manually.

Image Analysis

Fiji 1.0 (ImageJ2) (Schindelin et al., 2012) was used for image analysis. Fluorescence intensity across hPSCs (Figure 10D and 10E) was measured using the Plot Profile function. For each condition, 48 cells from 6 randomly selected fields were analysed. Fluorescence intensity (Figure 6A and 6B) was quantified using Cell Profiler Software.

Quantitative PCR

Total RNA was isolated using Total RNA Purification Kit (Norgen Biotek Cat. 37500), and complementary DNA (cDNA) was made from 500 ng using M-MLV Reverse Transcriptase (Invitrogen Cat. 28025-013) and dN6 primers. For real-time PCR SYBR Green Master mix (Bioline BIO-94020) was used. Primers are detailed in Table 3.

Three technical replicates were carried out for all quantitative PCR. GAPDH was used as endogenous control to normalize expression.

RNA-seq

For induction experiments, poly(A) mRNA was purified from total RNA using the Dynabeads "mRNA direct kit" (ThermoFisher, Cat. 61011). Quantity and quality of the starting mRNA were checked by Qubit and Agilent Bioanalyzer 2100 RNA pico chip. The template library was prepared using the "Ion Total RNA-Seq Kit v2" (ThermoFisher, Cat. 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 pM library was performed using a OneTouch 2 instrument (ThermoFisher, Cat. 4474778) with an Ion PI Template OT2 200 kit following the manufacturer's instructions (ThermoFisher, Cat. 4488318). The enrichment of template library was achieved using the Ion OneTouch ES enrichment system (ThermoFisher). Ion Proton sequencer and IPv2 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, reads that were not aligned in this step were realigned with bowtie2. Raw counts over the ensembl annotation release 75 were obtained with htseq-count v 0.6.0. Normalization and differential analysis were carried out using edgeR package (Robinson et al., 2010) and R (v3.0.0). Raw counts were normalized to obtain Counts Per Million mapped reads (CPM) and Reads Per Kilobase per Million mapped reads (RPKM). Only genes with a CPM greater than 1 in at least 2 samples were retained for differential analysis. Differences between batches were adjusted using an additive model. Genes were considered significantly up-regulated with a p-value less than or equal to 0.05 and a fold-change greater than or equal to 1.5.

For overexpression experiments, approximately 2 µg of total RNA were subjected to poly(A) selection, and libraries were prepared using the TruSeq RNA Sample Prep Kit (Illumina) following manufacturer's instructions. Sequencing was performed on the

Illumina NextSeq 500 platform. Reads were mapped to the *Homo sapiens* hg19 reference assembly using TopHat v2.1.1, and gene counts were computed using htseq-count (Anders et al., 2015). Differential expression analysis was performed using DESeq v2 (Anders and Huber, 2010). Genes with abs(Log2 fold-change) ≥ 1 and p-value < 0.01 were considered significant and defined as differentially expressed (Differentially Expressed Genes = DEG).

Gene Ontology (GO) terms for biological processes analysis of DEGs was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) database (Dennis et al., 2003) (https://david.ncifcrf.gov).

Boxplots and Scatterplots were made using TPM values exploiting ggpubr R package (v. 0.2), ggboxplot and ggscatter functions respectively.

Heatmaps were produced using TPM values with the pheatmap function from pheatmap R package (v.1.0.12, distance = 'correlation', scale = 'row') on selected markers.

Volcano plots were computed with Log2 fold-change and −log10 p-value using EnhancedVolcano function from EnhancedVolcano R package (v. 1.0.1).

Microarray

Public gene expression data of human embryonic stem cells treated with SB43 were downloaded from ArrayExpress (E-MEXP-1741). Differentially expressed genes were identified applying limma (Ritchie et al., 2015) on the RMA normalized gene expression matrix. To identify genes associated with TGF-beta inhibition, we compared the expression levels of human embryonic stem cells treated with SB43 or control cells and selected those probe sets with a fold change lower than or equal to -2 and an FDR lower than or equal to 0.05. Microarray analyses were performed in R (version 3.0.0).

ChIP-seq

ChIP-seq data of SMAD3 in BG03 embryonic stem cells were retrieved from GEO (GSE21614). We analysed the chromatin IP against Smad3 (GSM539548) and whole cell extract (WCE) in the same cell line (GSM539552). Raw reads were aligned using Bowtie (version 0.12.7); (Langmead et al., 2009) to build version hg19 of the human genome retaining only uniquely mapped reads. Redundant reads were removed using SAMtools. MACS2 version 2.0.10 (Zhang et al., 2008) was used to call peaks for SMAD3 using WCE ChIP-seq as control sample and setting the bandwidth equal to the estimated sonication fragment size (131 bp) and the p-value cutoff at 0.01. Only peaks with a pileup height greater than 5 were kept for further analysis. Each peak was assigned to the nearest TSS in a window of 100 Kb centered on the peak, considering only protein coding genes in GENCODE v16 annotation.

For identification of ZNF398 targets, we performed Chromatin Immunoprecipitation as previously described (Kim et al., 2009b; Krepelova et al., 2014) in two independent hESC lines (H9 and BG01V). ChIP-seq library was prepared with approximately 5 ng of immunoprecipitated DNA were used as input for the NEBNext® ChIP-Seq Library Prep kit, following manufacturer's instructions. Sequencing was performed on the Illumina NextSeq 500 platform. Reads were mapped to the *Homo sapiens* hg19 reference assembly using Bowtie v1.2.2 (Langmead et al., 2009), keeping only uniquely mapped reads. Reads (75 bp) were bioinformatically 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 v2.1.1 (Zhang et al., 2008), selecting only peaks with q-value < 0.05. A non-redundant set of common peaks between the two ZNF398 ChIP-seq replicates was generated using the *intersectBed* utility from BEDTools (Quinlan and Hall, 2010). For motif discovery, peaks were resized to ±200 bp surrounding their center and motif discovery was performed using MEME v4.10.1 (Bailey et al., 2009). 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

H3K4me1), GSE76084 (H3K27me3, H3K36me3, H3K4me3, H3K9ac, SOX2), GSE118325 (H3K9me3), GSE73725 (NANOG). Data for POU5F1 and EP300 was instead obtained from the ENCODE database (https://www.encodeproject.org/). All samples were analyzed as stated above. Correlations between genomic occupancy profiles were computed using the *multiBamSummary* and *plotCorrelation* utilities from deepTools v2.2.4 (Ramírez et al., 2014). Heatmaps of peak densities around ZNF398 peaks centers were generated using in-house developed scripts.

Statistics and reproducibility

For each dataset, sample size *n* refers to experimental replicates and is represented by the number of dots in the plots or stated in the figure legends. Normality distribution was not assumed and p-values were calculated using the non-parametric unpaired two-tailed Mann-Whitney *U* test with the exception of induction experiments (Figure 3D and 3E) 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, indicated as their numerical values or as asterisks (* P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001). R software was used for statistical analysis (R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/).

All error bars indicate the standard error of the mean (s.e.m.). All key experiments were repeated between two and five times independently, as indicated. Experiments of candidates functional validation were repeated using 3 different hPSC lines. All qPCR experiments were performed with three technical replicates.

Data availability

RNA-seq and ChiP-seq data of this study have been deposited on Gene Expression Omnibus (GEO) database, under the accession GSE133630.

For the identification of TGF-beta transcriptional targets, we used available SMAD3 ChIP-seq data from (Mullen et al., 2011) (Accession Number GSE21621), microarray

data from (Vallier et al., 2009) (Accession Number E-MEXP-1741) and RNA-seq data of H9 from (Rada-Iglesias et al., 2011), (Accession Number GSE24447), see Figure 3C. All plasmids generated in this study will be deposited to Addgene upon publication of the work.

Table 1: siRNA sequence information

Name	Product code	Sequence	Company
siCONTROL	SI03650318	UUCUCCGAACGUGUCACGU	Qiagen
siGFP	Custom	GCAAGCTGACCCTGAAGTTCA	Sigma
siNANOG 1	SI04341400	CACCCAATCCTGGAACAATCA	Qiagen
siNANOG 2	SI00654738	TTGGTTTAAGTTCAAATGAAT	Qiagen
siNANOG 3	SI05785122	AGGCCTTAATGTAATACAGCA	Qiagen
siNANOG 4	SI05785108	CTGCATGCAGTTCCAGCCAAA	Qiagen
siNANOG 5	SI05785115	CTCCATGGATCTGCTTATTCA	Qiagen
siZNF398 1	Custom	GAAGGGCAACTACGAGTCTCT	Sigma
siZNF398 2	SI05785164	CCGCAAGCACCACCTAATGAA	Qiagen
siZNF398 3	SI05785157	CAGGTGCGATACTAAGCCTCA	Qiagen
siZNF398 4	SI05785178	CTCCATGGATTATGCTATAAA	Qiagen
siZNF398 5	Custom	GAATCTCAGCCAAGACATGTT	Sigma

Table 2: Antibodies details

Antibody	Dilution	Product code	Company	Antibody validation
NANOG	1:100	D73G4	Cell Signaling TECHNOLOGY	Staining colocalizes with nucleus in hPSCs in presence of TGF-beta and decrease upon TGF-beta deprivation
POU5F1	1:300	sc-5279	Santa Cruz	Staining colocalizes with nucleus in hPSCs in presence of TGF-beta and decrease upon TGF-beta deprivation
β-CATENIN	1:100	sc-7963	Santa Cruz	Staining colocalizes with cytoskeleton in epithelial cells (Previuosly validated by Azzolin L. et al., 2014)
VIMENTIN	1:100	sc-7557	Santa Cruz	Staining colocalizes with cytoskeleton in mesenchymal cells (hPSCs treated with SB43, so in absence of TGF-beta)
ECAD	1:100	610181	BD	Staining colocalizes with cytoskeleton in epithelial cells (hPSCs, in presence of TGF-beta)
SMAD2/3	1:50	610843	BD	Staining colocalizes with cytoplasm in absence of TGF-beta and it shifts in the nucleus upon TGF-beta stimulation
Alexa Fluor 488 Phalloidin	1:200	A12379	ThermoFisher	Detects filamentous cytoplasmic structures that are lost upon treatment of cells with

				F-actin inhibitory drugs.
Donkey anti-Rabbit IgG (H+L) Secondary Antibody Alexa Fluor 488	1:500	A-21206	ThermoFisher	Internally validated
Goat anti-Rabbit IgG (H+L) Secondary Antibody Alexa Fluor 568	1:500	A-11036	ThermoFisher	Internally validated
Donkey anti-Mouse IgG (H+L) Secondary Antibody Alexa Fluor 488	1:500	A-21202	ThermoFisher	Internally validated
Donkey anti-Mouse IgG (H+L) Secondary Antibody Alexa Fluor 647	1:500	A-31571	ThermoFisher	Internally validated
Donkey anti-Mouse IgG (H+L) Secondary Antibody Alexa Fluor 568	1:500	A-10037	ThermoFisher	Internally validated

Table 3: qPCR primers details

Gene	Forward	Reverse
3'UTR	GTAATCATGGTCATAGCTGTTTCCT	CCAGGCTTTACACTTTATGCTTC
AFF4	TCTCAGTCTCAGAAACGGTCC	GGCTACTGCTCCCACTATTGTT
BCOR	CTCAGGGGCCACGAAACT	ATGGTGACAATTTCCACGTTT
CCRN4L	TCTCGCCAAGACACTGAACAG	GGCCCTGCATTCCTCAAGAAG
CXX5	GCTGCTCTGGAGAAGGTGAT	GACCACACGAGCAGTGACAT
DPPA4	TCTGGTGTCAGGTGGTGTGT	TCCCTTCTTGCTTTTCTGGA
ECAD	GGTCTGTCATGGAAGGTGCT	TCAGGATCTTGGCTGAGGAT
ENC1	TTGGAGCTGCTGCTTGACTA	TCCCGGATGTCTTGAAACTC
EPCAM	TGGACATAGCTGATGTGGCTTA	CCAGGATCCAGATCCAGTTG
ESRP1	CTTCCAACCCCTCCCATTAT	ATTGTGGCTGCATAGGGAAG
ETS2	CTGGGCATTCCAAAGAACCC	CCAGACTGAACTCATTGGTGG
ETV1	GGCCCCAGGCAGTTTTATGAT	GATCCTCGCCGTTGGTATGT
EZR	TGCAATCCAGCCAAATACAA	TTATCCAGCTTCAGCCAGGT
FLK1/KDR	GGCGGCACGAAATATCCTCT	GGAGGCGAGCATCTCCTTTT
FOXO1	TGATAACTGGAGTACATTTCGCC	CGGTCATAATGGGTGAGAGTCT
GAPDH	CGAGATCCCTCCAAAATCAA	GGCAGAGATGATGACCCTTT
GATA6	GCAAAAATACTTCCCCCACA	TCTCCCGCACCAGTCATC
GBX2	GTTCCACTGCAAAAAGTACCTCT	GGGACGACGATCTTAGGGTTC
ID1	ATCGCATCTTGTGTCGCTGA	GCCGATCGGTCTTGTTCTCC
KLF10	GCCAACCATGCTCAACTTCG	TGCAGTTTTGTTCCAGGAATACAT
KLF7	CTCATGGGAGGGATGTGAGT	ACCTGGAAAAACACCTGTCG
LEFTY1	GGACCTTGGGGACTATGGAG	ATCCCCTGCAGGTCAATGTA
LIN28A	CTGTAAGTGGTTCAACGTGCG	CCATGTGCAGCTTACTCTGGT
LIN28B	CCTCCTCAGCCAAAGAAGTG	TGGGATTCTGCTTCCTGTCT
MDF1	CTGGAGATCTGCATGGAGTG	AGAGGGACCCTGCTGGAAT
MYC	AGCGACTCTGAGGAGGAACA	GCTGTGAGGAGGTTTGCTGT
NANOG	TTTGTGGGCCTGAAGAAAACT	AGGGCTGTCCTGAATAAGCAG
NCAD	CGTGGTCAAACCAATCGAC	AACAGACACGGTTGCAGTTG
NNAT	CCCACCATGGAAATCAAAAC	ACCACCCTCCTTCCTCAACT
OCT4	GTGGAGGAAGCTGACAACAA	ATTCTCCAGGTTGCCTCTCA
OTX2	CAAAGTGAGACCTGCCAAAAAGA	TGGACAAGGGATCTGACAGTG
PAX3	TCCACAAGCTGTGTCAGATCC	GCGTTGGAAGGAATCGTGCT
PAX6	TGGGCAGGTATTACGAGACTG	ACTCCCGCTTATACTGGGCTA
PRDM14	GAGCCTTCAGGTCACAGAGC	TCCACACAGGGGGTGTACTT
S100A4	AGCTTCTTGGGGAAAAGGAC	TCTTGGAAGTCCACCTCGTT
SIP1	CGCTTGACATCACTGAAGGA	CTTGCCACACTCTGTGCATT
SIX3	CCCCACCTGCATGACGATTT	TCTCCTCGTGGTGGTGATTG
SKIL	TATGCAGGACAGTTGGCAGA	TTCTGTCTTGCTTCCCGTTC
SMAD7	CAAGAGGCTGTGTTGCTGTG	GGGAGACTCTAGTTCGCAGA
SOX17	ACGCCGAGTTGAGCAAGA	TCTGCCTCCTCCACGAAG

T	TATGAGCCTCGAATCCACATAGT	CCTCGTTCTGATAAGCAGTCAC
TBL1XR1	GGCAGAGCAACAACACCTTT	TGGGTCCCATTTGATAGCAT
TEAD1	ATGCCAACCATTCTTACAGTGAC	ACAGTTCCTTTAAGCCACCTTTC
TLE3	TTGAGCCGATACGACAGTGA	CTTGTCCAGCCCATTTTCAG
UTF1	CTCCCAGCGAACCAGACG	GGAGGCGTCCGCAGACTT
VIM	CTCCACGAAGAGGAAATCCA	GTGAGGTCAGGCTTGGAAAC
ZFP42	GCCTTATGTGATGGCTATGTGT	ACCCCTTATGACGCATTCTATGT
ZNF398	TGGCAAGAATCTCAGCCAAGA	GTGGAGTAAAGTGCTTAGGGC

REFERENCES

Adachi, K., Suemori, H., Yasuda, S.-Y., Nakatsuji, N., and Kawase, E. (2010). Role of SOX2 in maintaining pluripotency of human embryonic stem cells. Genes Cells Devoted Mol. Cell. Mech. *15*, 455–470.

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

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

Babaie, Y., Herwig, R., Greber, B., Brink, T.C., Wruck, W., Groth, D., Lehrach, H., Burdon, T., and Adjaye, J. (2007). Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. Stem Cells Dayt. Ohio *25*, 500–510.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202-208.

Bedzhov, I., and Zernicka-Goetz, M. (2014). Self-Organizing Properties of Mouse Pluripotent Cells Initiate Morphogenesis upon Implantation. Cell *156*, 1032–1044.

Boroviak, T., and Nichols, J. (2017). Primate embryogenesis predicts the hallmarks of human naïve pluripotency. Development *144*, 175–186.

Bouma, M.J., van Iterson, M., Janssen, B., Mummery, C.L., Salvatori, D.C.F., and Freund, C. (2017). Differentiation-Defective Human Induced Pluripotent Stem Cells Reveal Strengths and Limitations of the Teratoma Assay and In Vitro Pluripotency Assays. Stem Cell Rep. *8*, 1340–1353.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva De Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. *27*, 275–280.

Chappell, J., and Dalton, S. (2013). Roles for MYC in the Establishment and Maintenance of Pluripotency. Cold Spring Harb. Perspect. Med. *3*, a014381.

Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E., et al. (2011). Chemically defined conditions for human iPSC derivation and culture. Nat. Methods *8*, 424–429.

Chen, K.G., Mallon, B.S., McKay, R.D.G., and Robey, P.G. (2014). Human pluripotent stem cell culture: Considerations for maintenance, expansion, and therapeutics. Cell Stem Cell *14*, 13–26.

Chen, S., Choo, A., Chin, A., and Oh, S.K.W. (2006). TGF-β2 allows pluripotent human embryonic stem cell proliferation on E6/E7 immortalized mouse embryonic fibroblasts. J. Biotechnol.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e20.

Cliff, T.S., Wu, T., Boward, B.R., Yin, A., Yin, H., Glushka, J.N., Prestegaard, J.H., and Dalton, S. (2017). MYC Controls Human Pluripotent Stem Cell Fate Decisions through Regulation of Metabolic Flux. Stem Cell 1–47.

Conroy, A.T., Sharma, M., Holtz, A.E., Wu, C., Sun, Z., and Weigel, R.J. (2002). A novel zinc finger transcription factor with two isoforms that are differentially repressed by estrogen receptor-alpha. J. Biol. Chem. *277*, 9326–9334.

Daley, G.Q., Lensch, M.W., Jaenisch, R., Meissner, A., Plath, K., and Yamanaka, S. (2009). Broader implications of defining standards for the pluripotency of iPSCs. Cell Stem Cell *4*, 200–201; author reply 202.

Darr, H., Mayshar, Y., and Benvenisty, N. (2006). Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. Dev. Camb. Engl. *133*, 1193–1201.

Deglincerti, A., Croft, G.F., Pietila, L.N., Zernicka-Goetz, M., Siggia, E.D., and Brivanlou, A.H. (2016). Self-organization of the *in vitro* attached human embryo. Nature *533*, 251–254.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. *4*, R60.

Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J.M. (1998). Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. EMBO J. *17*, 3091–3100.

Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science *344*, 1156–1160.

Eastham, A.M., Spencer, H., Soncin, F., Ritson, S., Merry, C.L.R., Stern, P.L., and Ward, C.M. (2007). Epithelial-Mesenchymal Transition Events during Human Embryonic Stem Cell Differentiation. Cancer Res. *67*, 11254–11262.

Ferraris, L., Stewart, A.P., Kang, J., DeSimone, A.M., Gemberling, M., Tantin, D., and Fairbrother, W.G. (2011). Combinatorial binding of transcription factors in the pluripotency control regions of the genome. Genome Res. *21*, 1055–1064.

Fong, H., Hohenstein, K.A., and Donovan, P.J. (2008). Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. Stem Cells Dayt. Ohio *26*, 1931–1938.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. *85*, 348–362.

Gagliano, O., Luni, C., Qin, W., Bertin, E., Torchio, E., Galvanin, S., Urciuolo, A., and Elvassore, N. (2019). Microfluidic reprogramming to pluripotency of human somatic cells. Nat. Protoc. *14*, 722–737.

Giulitti, S., Pellegrini, M., Zorzan, I., Martini, P., Gagliano, O., Mutarelli, M., Ziller, M.J., Cacchiarelli, D., Romualdi, C., Elvassore, N., et al. (2018). Direct generation of human naive induced pluripotent stem cells from somatic cells in microfluidics. Nat. Cell Biol.

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

Hackett, J.A., Huang, Y., Günesdogan, U., Gretarsson, K.A., Kobayashi, T., and Surani, M.A. (2018). Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. Nat. Commun. *9*.

Høffding, M.K., and Hyttel, P. (2015). Ultrastructural visualization of the Mesenchymal-to-Epithelial Transition during reprogramming of human fibroblasts to induced pluripotent stem cells. Stem Cell Res. *14*, 39–53.

Höpfl, G., Gassmann, M., and Desbaillets, I. (2004). Differentiating Embryonic Stem Cells into Embryoid Bodies. In Germ Cell Protocols, (New Jersey: Humana Press), pp. 079–098.

Hyslop, L., Stojkovic, M., Armstrong, L., Walter, T., Stojkovic, P., Przyborski, S., Herbert, M., Murdoch, A., Strachan, T., and Lako, M. (2005). Downregulation of NANOG Induces Differentiation of Human Embryonic Stem Cells to Extraembryonic Lineages. STEM CELLS *23*, 1035–1043

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

Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., and Benvenisty, N. (2000). Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. Mol. Med. Camb. Mass *6*, 88–95.

Jeter, C.R., Yang, T., Wang, J., Chao, H.-P., and Tang, D.G. (2015). Concise Review: NANOG in Cancer Stem Cells and Tumor Development: An Update and Outstanding Questions. STEM CELLS *33*, 2381–2390.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat. Cell Biol. *10*, 353–360.

Johansson, B.M., and Wiles, M.V. (1995). Evidence for Involvement of Activin A and Bone Morphogenetic Protein 4 in Mammalian Mesoderm and Hematopoietic Development. Mol. Cell. Biol. *15*, 141–151.

Kim, D., Kim, C.H., Moon, J.I., Chung, Y.G., Chang, M.Y., Han, B.S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., et al. (2009a). Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins. Cell Stem Cell *4*, 472–476.

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

Krepelova, A., Neri, F., Maldotti, M., Rapelli, S., and Oliviero, S. (2014). Myc and Max genome-wide binding sites analysis links the Myc regulatory network with the polycomb and the core pluripotency networks in mouse embryonic stem cells. PLoS ONE *9*, 1–12.

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

Levenstein, M.E., Ludwig, T.E., Xu, R.-H., Llanas, R.A., VanDenHeuvel-Kramer, K., Manning, D., and Thomson, J.A. (2006). Basic Fibroblast Growth Factor Support of Human Embryonic Stem Cell Self-Renewal. Stem Cells *24*, 568–574.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. Cell Stem Cell *7*, 51–63.

Löser, P., Schirm, J., Guhr, A., Wobus, A.M., and Kurtz, A. (2010). Human Embryonic Stem Cell Lines and Their Use in International Research. STEM CELLS *28*, 240–246.

Lu, T.-Y., Lu, R.-M., Liao, M.-Y., Yu, J., Chung, C.-H., Kao, C.-F., and Wu, H.-C. (2010). Epithelial Cell Adhesion Molecule Regulation Is Associated with the Maintenance of the Undifferentiated Phenotype of Human Embryonic Stem Cells. J. Biol. Chem. *285*, 8719–8732.

Ludwig, T.E., Levenstein, M.E., Jones, J.M., Berggren, W.T., Mitchen, E.R., Frane, J.L., Crandall, L.J., Daigh, C.A., Conard, K.R., Piekarczyk, M.S., et al. (2006). Derivation of human embryonic stem cells in defined conditions. Nat. Biotechnol. *24*, 185–187.

Luni, C., Giulitti, S., Serena, E., Ferrari, L., Zambon, A., Gagliano, O., Giobbe, G.G., Michielin, F., Knöbel, S., Bosio, A., et al. (2016). High-efficiency cellular reprogramming with microfluidics. Nat. Methods *13*, 446–452.

Martello, G., and Smith, A. (2014). The nature of embryonic stem cells. Annu. Rev. Cell Dev. Biol. *30*, 647–675.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell *11*, 491–504.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J. *32*, 2561–2574.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U. S. A. *78*, 7634–7638.

Massagué, J., Seoane, J., and Wotton, D. (2005). Smad transcription factors. Genes Dev. *19*, 2783–2810.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat. Cell Biol. *9*, 625–635.

McConnell, B.B., and Yang, V.W. (2010). Mammalian Krüppel-Like Factors in Health and Diseases. Physiol. Rev. *90*, 1337–1381.

Mullen, A.C., Orlando, D.A., Newman, J.J., Lovén, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., Dekoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF-β signaling. Cell *147*, 565–576.

Muñoz-Sanjuán, I., and Brivanlou, A.H. (2002). Neural induction, the default model and embryonic stem cells. Nat. Rev. Neurosci. *3*, 271–280.

Nagy, A., Gócza, E., Diaz, E.M., Prideaux, V.R., Iványi, E., Markkula, M., and Rossant, J. (1990). Embryonic stem cells alone are able to support fetal development in the mouse. Dev. Camb. Engl. *110*, 815–821.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. *26*, 101–106.

Nichols, J., and Smith, A. (2009). Naive and Primed Pluripotent States. Cell Stem Cell *4*, 487–492.

Nichols, J., and Smith, A. (2011). The origin and identity of embryonic stem cells. Development *138*, 3–8.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379–391.

Nishimura, K., Sano, M., Ohtaka, M., Furuta, B., Umemura, Y., Nakajima, Y., Ikehara, Y., Kobayashi, T., Segawa, H., Takayasu, S., et al. (2011). Development of Defective and Persistent Sendai Virus Vector A UNIQUE GENE DELIVERY/EXPRESSION SYSTEM IDEAL FOR CELL REPROGRAMMING. J. Biol. Chem. *286*, 4760–4771.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat. Genet. *24*, 372.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K.I., et al. (2011). A more efficient method to generate integration-free human iPS cells. Nat. Methods *8*, 409–412.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315.

Pei, D., Shu, X., Gassama-Diagne, A., and Thiery, J.P. (2019). Mesenchymal–epithelial transition in development and reprogramming. Nat. Cell Biol. *21*, 44–53.

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

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

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

Rodriguez-Boulan, E., and Macara, I.G. (2014). Organization and execution of the epithelial polarity programme. Nat. Rev. Mol. Cell Biol. *15*, 225–242.

Ross, S., and Hill, C.S. (2008). How the Smads regulate transcription. Int. J. Biochem. Cell Biol. *40*, 383–408.

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

Sánchez-Castillo, M., Ruau, D., Wilkinson, A.C., Ng, F.S.L., Hannah, R., Diamanti, E., Lombard, P., Wilson, N.K., and Göttgens, B. (2015). CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. Nucleic Acids Res. *43*, D1117–23.

Schier, A.F. (2003). Nodal signaling in vertebrate development. Annu. Rev. Cell Dev. Biol. *19*, 589–621.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schlaeger, T.M., Daheron, L., Brickler, T.R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., et al. (2015). A comparison of non-integrating reprogramming methods. Nat. Biotechnol. *33*, 58–63.

Scholer, H.R., Dressler, G.R., Rohdewohid, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. Trends Genet. *6*, 314.

Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D.A., and Benvenisty, N. (2000). Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. Proc. Natl. Acad. Sci. *97*, 11307–11312.

Shahbazi, M.N., and Zernicka-Goetz, M. (2018). Deconstructing and reconstructing the mouse and human early embryo. Nat. Cell Biol. 1–10.

Shahbazi, M.N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., Fogarty, N.M.E., Campbell, A., Devito, L.G., Ilic, D., et al. (2016). Self-organization of the human embryo in the absence of maternal tissues. Nat. Cell Biol. *18*, 700–708.

Shahbazi, M.N., Siggia, E.D., and Zernicka-Goetz, M. (2019). Self-organization of stem cells into embryos: A window on early mammalian development. Science *364*, 948–951.

Shao, Y., Taniguchi, K., Gurdziel, K., Townshend, R.F., Xue, X., Yong, K.M.A., Sang, J., Spence, J.R., Gumucio, D.L., and Fu, J. (2016). Self-organized amniogenesis by human pluripotent stem cells in a biomimetic implantation-like niche. Nat. Mater. *16*, 419–425.

Shao, Y., Taniguchi, K., Townshend, R.F., Miki, T., Gumucio, D.L., and Fu, J. (2017). A pluripotent stem cell-based model for post-implantation human amniotic sac development. Nat. Commun. 1–15.

Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685–700.

Shi, Y., Wang, Y.F., Jayaraman, L., Yang, H., Massagué, J., and Pavletich, N.P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. Cell *94*, 585–594.

Shu, X., and Pei, D. (2014). The function and regulation of mesenchymal-to-epithelial transition in somatic cell reprogramming. Curr. Opin. Genet. Dev. *28*, 32–37.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. Cell *138*, 722–737.

Simunovic, M., Metzger, J.J., Etoc, F., Yoney, A., Ruzo, A., Martyn, I., Croft, G., You, D.S., Brivanlou, A.H., and Siggia, E.D. (2019). A 3D model of a human epiblast reveals BMP4-driven symmetry breaking. Nat. Cell Biol. *21*, 900–910.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell *158*, 1254–1269.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K., and Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proc. Natl. Acad. Sci. *110*, 12172–12179.

Taniguchi, K., Shao, Y., Townshend, R.F., Tsai, Y.-H., DeLong, C.J., Lopez, S.A., Gayen, S., Freddo, A.M., Chue, D.J., Thomas, D.J., et al. (2015). Lumen Formation Is an Intrinsic Property of Isolated Human Pluripotent Stem Cells. Stem Cell Rep. *5*, 954–962.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Thomson, J.A. (1998). Embryonic Stem Cell Lines Derived from Human Blastocysts. Science *282*, 1145–1147.

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

Vallier, L., Reynolds, D., and Pedersen, R.A. (2004). Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. Dev. Biol. *275*, 403–421.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W.B., Cho, C.H.H., Martinez, A., Rugg-Gunn, P., et al. (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339–1349.

Villodre, E.S., Kipper, F.C., Pereira, M.B., and Lenz, G. (2016). Roles of OCT4 in tumorigenesis, cancer therapy resistance and prognosis. Cancer Treat. Rev. *51*, 1–9.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell *10*, 440–454.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.-H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., et al. (2010). Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. Cell Stem Cell *7*, 618–630.

Warzecha, C.C., Jiang, P., Amirikian, K., Dittmar, K.A., Lu, H., Shen, S., Guo, W., Xing, Y., and Carstens, R.P. (2010). An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. EMBO J. *29*, 3286–3300.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., et al. (2009). PiggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766–770.

Xu, R.-H., Sampsell-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R., et al. (2008). NANOG Is a Direct Target of TGFβ/Activin-Mediated SMAD Signaling in Human ESCs. Cell Stem Cell *3*, 196–206.

Yilmaz, A., Peretz, M., Aharony, A., Sagi, I., and Benvenisty, N. (2018). Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. Nat. Cell Biol. *20*, 610–619.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science *318*, 1917–1920.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2009). Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. Science *324*, 797–801.

Yusa, K., Rad, R., Takeda, J., and Bradley, A. (2009). Generation of transgene-free induced pluripotent mouse stem cells by the *piggyBac* transposon. Nat. Methods *6*, 363–369.

Zaehres, H., Lensch, M.W., Daheron, L., Stewart, S.A., Itskovitz-Eldor, J., and Daley, G.Q. (2005). High-Efficiency RNA Interference in Human Embryonic Stem Cells. STEM CELLS *23*, 299–305.

Zawel, L., Dai, J.L., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B., and Kern, S.E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. Mol. Cell *1*, 611–617.

Zhang, J., Ratanasirintrawoot, S., Chandrasekaran, S., Wu, Z., Ficarro, S.B., Yu, C., Ross, C.A., Cacchiarelli, D., Xia, Q., Seligson, M., et al. (2016). LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. Stem Cell *19*, 66–80.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M., Gumucio, D.L., et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. Nature *573*, 421–425.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., et al. (2009). Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. Cell Stem Cell *4*, 381–384.