Whsc1 links pluripotency exit with mesendoderm specification

Whsc1 links pluripotency exit with mesendoderm specification

Tian V. Tian[1,2], Bruno Di Stefano[1,2,3], Johanna Goldmann[1,2,4], José Luis Sardina[1,2], Alessandro Dasti[1,2], Carolina Segura Morales[1,2], Luisa De Andres Aguayo[1,2], Antonio Gomez Moruno[1,2] and Thomas Graf[1,2]

1. Gene Regulation, Stem Cells and Cancer Program, Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Carrer Dr Aiguader 88, 08003 Barcelona, Spain

2. Universitat Pompeu Fabra (UPF), Barcelona, Spain

3. Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA.

4. The Whitehead Institute, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

Correspondence: thomas.graf@crg.eu, tian.tian@crg.eu

Running title: Whsc1 links pluripotency exit and differentiation

Keywords: Whsc1, chromatin related factor, embryonic stem cells, pluripotency exit, and germ layer specification

Abstract

How pluripotent stem cells differentiate into the main germ layers is a key question of developmental biology. Here we show that the chromatin-related factor Whsc1 has a dual role in pluripotency exit and germ layer specification of embryonic stem cells (ESCs). Upon induction of differentiation, a proportion of Whsc1-depleted ESCs remain entrapped in a pluripotent state and fail to form mesoderm and endoderm, although they are still capable of generating neuroectoderm. These functions of Whsc1 are independent of its methyltransferase activity. Whsc1 binds to enhancers of the mesendodermal regulators *Gata4*, *Brachyury*, *Gata6* and *Foxa2* together with Brd4, and activates the genes' expression. Depleting each of these regulators also delays pluripotency exit, suggesting that they mediate the effects observed with Whsc1. Hence, up-regulation of mesendoderm instructive transcription factors is required for the timely egress from pluripotency. Our data suggest that the silencing of the pluripotency regulatory network and the activation of lineage-restricted networks are tightly interconnected.

Embryonic stem cells (ESCs) are an excellent model to study the generation of the three main germ layers (endoderm, mesoderm and ectoderm) and thus all bodily cell-types [1]. The first event during the transition of ESCs into differentiated cells is the exit from pluripotency, which includes the silencing of the core pluripotency transcription factors (TFs) Oct4, Sox2 and Nanog [2-4]. This process can either be externally initiated by deprivation of self-renewal signals [5] or by exposure of the cells to differentiation-inducing cues [6]. Subsequently, germ layer-instructive TFs become up-regulated and activate markers that define the various lineages [4]. These factors comprise Brachyury (T, Bra) and Gata4 for mesoderm, Gata6, Gata4 and Foxa2 for endoderm and Pax6 and Sox1 for neuroectoderm [3, 4, 6, 7]. Screens with ESC reporter lines and haploid ESCs have identified several regulators that are required for pluripotency exit. These include several TFs and RNA binding proteins, all of which were shown to affect members of the pluripotency network [8-11]. However, alternative mechanisms contributed by yet to be discovered factors still might exist.

Besides TFs and noncoding RNAs, chromatin-related factors (CRFs) play an important role in ESC differentiation, which involves a progressive transition from a relatively open chromatin state to a more compact one [12]. Several CRFs are broadly involved in ESC differentiation, such as Polycomb group proteins that are required for the repression of pluripotency associated and lineage inappropriate genes [13]. Others have been shown to act in a more restricted fashion, as exemplified by the requirement of Setd2 for endodermal [14], Mel18 for mesodermal [15] and Zrf1 for neuroectodermal differentiation [16].

Here we report that Whsc1, a CRF up-regulated during embryoid body differentiation, is required for efficient pluripotency exit of ESCs and for their specific differentiation towards mesendoderm. However, unlike the factors reported before, Whsc1 does not impair the pluripotency network after induction of differentiation but activates enhancers of key mesendoderm TFs. Depletion of these downstream mesendoderm TFs also delays exit from pluripotency, explaining the effects seen with Whsc1. Our data uncover a new role for Whsc1 and establish an intimate connection between regulatory networks that define pluripotent cells on the one hand and differentiated cells on the other.

Whsc1 is required for efficient exit from pluripotency

To search for CRFs that play a role in pluripotency exit and subsequent ESC differentiation, we performed an *in silico* screen by analysing datasets recording

transcriptome changes during the differentiation of mouse ESCs into embryoid bodies (EBs). We compiled a list of 653 genes encoding histone writers and readers, histone variants, nucleosome positioning proteins and a selection of well-annotated long non-coding RNAs (Supplementary Table 1 and Supplementary Fig. 1a). We focused on three candidates that consistently scored among the most up-regulated genes. These were *Cbx4,* encoding a Polycomb group (PcG) protein that has been shown to orchestrate ESC differentiation [17], *L3mbtl3*, a putative PcG family member [18] and *Whsc1*, encoding a SET-domain methyltransferase associated with Wolf-Hirschhorn syndrome [19].

To study whether these candidates are required for pluripotency exit, we tested the effects of knockdowns in a pluripotency reporter ESC line induced to differentiate. This line, which contains a destabilized version of GFP knocked into the *Rex1* locus [9], was induced to differentiate by transfer into medium lacking 2i and LIF [20] but containing with N2B27 supplements as well as Activin A and foetal bovine serum (FBS), which favour the formation of neuroectodermal and mesendodermal cells [21, 22]. Upon induction, the reporter line rapidly became GFP negative (Fig. 1a-c) as the loss of Rex1 expression is a highly sensitive readout for exit from the pluripotent state [23, 24]. Accordingly, the cells showed a dramatic down-regulation of pluripotency genes and an increased expression of mesendoderm and neuroectoderm specific genes (Supplementary Fig. 1b). Two independent small hairpin RNA (shRNA) lentiviral constructs were used to knock down the expression of each of the three candidates. Control ESCs expressing a scrambled shRNA control construct (shScr) became essentially GFP negative within 72 hours after induction, and similar findings were obtained with cells expressing the shCbx4 and shL3mbtl3 constructs (Supplementary Fig. 1c). In contrast, 25-35% of ESCs transduced with the shWhsc1 constructs retained GFP expression (Fig. 1b, c). The Whsc1 knockdown constructs resulted in > 70% decrease of Whsc1 mRNA and dramatically reduced the expression of both protein isoforms, with shWhsc1.4 being more effective than shWhsc1.2 (Fig. 1d). Interestingly, shWhsc1 ESCs retained elevated expression levels of the pluripotency genes *Rex1, Pou5f1 (Oct4), Nanog, Lin28a* and *Esrrb*, with shWhsc1.4 cells showing a stronger effect (Fig. 1e). In addition, Whsc1-depleted cells, when replated after induced differentiation, yielded large numbers of alkaline phosphatase positive (AP) colonies even after 2 consecutive rounds of treatment, whereas no colonies could be recovered from control cells already after one round (Supplementary Fig. 1d). Of note, depletion of Whsc1 in ESCs did not result in either an up-regulation of pluripotency

factor expression nor in an alteration of the cells' growth kinetics (Supplementary Fig. 1e-g), ruling out more indirect effects on pluripotency exit.

In summary, using an shRNA-based loss-of-function approach, we tested the effects of three candidates from an *in silico* screen for CRFs on the exit from pluripotency. Of these Whsc1 depletion was found to impair the efficient down-regulation of Rex1 and other pluripotency markers, indicating that Whsc1 is required for efficient pluripotency exit.

Depletion of Whsc1 impairs mesendoderm differentiation and down-regulation of pluripotency markers *in vitro* and *in vivo*

To further evaluate the effect of Whsc1 depletion on the differentiation capacity of ESCs, we examined embryoid bodies (EBs) derived from knockdown ESCs (Fig. 2a). Whsc1-depleted ESCs generated smaller EBs compared to those from control ESCs (Fig. 2b) and expressed lower levels of Whsc1 mRNA and protein and retained elevated levels of *Rex1* and *Pou5f1* (Fig. 2c,d), as expected. Surprisingly, Whsc1-depleted EBs also showed a strongly reduced expression of the mesendoderm regulators *Gata4*, *Flk1*, *Bra*, *Gata6*, *Foxa2* and *Sox17* (Fig. 2c) while neuroectodermal gene expression (*Sox1*, *Pax6* and *Nes*) remained unaffected (Fig. 2c). In addition, when day 17 shWhsc1 EBs were disassociated and replated on mouse embryo fibroblasts in medium with 2i and LIF they formed large numbers of AP+ colonies, with control cells yielding few such colonies (Fig. 2e), suggesting that they retained pluripotent cells. To probe their differentiation potential *in vivo*, we injected Whsc1 knockdown and control ESCs into immunodeficient mice, both of which generated teratomas. Control cell derived teratomas contained tissues from all three germ layers, such as muscle, cartilage, gut epithelium, neuro-epithelium. In contrast, Whsc1-depleted teratomas were mainly composed of neuro-epithelium and poorly differentiated cells (Supplementary Fig. 2a). Supporting these observations, qRT-PCR analyses showed that shWhsc1-derived teratomas exhibited retention of pluripotency markers and significantly reduced levels of mesendoderm markers while retaining levels of neuroectoderm markers comparable to the controls (Supplementary Fig. 2b).

These results suggest that Whsc1 links pluripotency exit and the specification towards mesendoderm. However, since both EBs and teratomas comprise a complex mixture of cells, we decided to test the effects of Whsc1 knockdown on the directed differentiation of ESCs towards mesoderm, endoderm and ectoderm, respectively. First we induced cardiac mesoderm differentiation [17] and found that Whsc1-depleted

cells generated significantly fewer beating colonies than control cells (Supplementary Fig. 3a, Supplementary Movie 1, 2). Supporting this observation, 10 days after induction Whsc1-depleted cells showed reduced expression of the cardiac regulators *Mef2c* and *Nkx2.5* while retaining *Rex1* and *Pou5f1* expression (Supplementary Fig. 3b, c). Next, we examined the effects of shWhsc1 during definitive endoderm differentiation, using Eomes-GFP reporter ESCs [25] and transferring the cells into medium containing Activin A, Fgf4, heparin, PI-103 and CHIR [26]. Seven days later about 30% of control cells became GFP positive, whereas only 5% of Whsc1-depleted cells did so (Supplementary Fig. 3d,e). In line with these findings, shWhsc1-expressing cells only partially up-regulated the endoderm markers *Cxcr4, Cldn6, Foxa2, Gata6* and *Sox17* and inefficiently down-regulated the pluripotency regulators *Rex1*, *Pou5f1* and *Esrrb* (Supplementary Fig. 3f, g). Finally, we derived neural progenitor cells [27] from control and knockdown ESCs and found no significant differences in the expression of the neural markers *Pax6, Sox1* and *Nes* nor that of *Pou5f1* and *Esrrb* (Supplementary Fig. 3h).

Taken together, the results obtained with embryoid bodies and teratomas extend our previous observations obtained in monolayers, namely that the delay of pluripotency exit in Whsc1-depleted ESCs is coupled specifically to an impairment of mesendoderm differentiation. In addition, similar results were obtained using lineage specific differentiation protocols.

The SET-domain of Whsc1 is dispensable for efficient pluripotency exit and mesendoderm differentiation

Whsc1 requires the catalytic SET-domain on the C-terminus to di-methylate the lysine 36 on histone 3 (Supplementary Fig. 4a) [28, 29]. This function has been linked to the normal foetal heart and cartilage development in a murine model in which SET-domain of Whsc1 was genetically excised [30]. Surprisingly, however, the ESCs derived from these mice (hereafter referred to as ΔSET) behaved like WT ESCs in our assays. Thus, after induction of differentiation in monolayers, both WT and ΔSET cells rapidly down-regulated pluripotency factors and up-regulated Whsc1 (Supplementary Fig. 4c). Likewise, EBs derived from WT and ΔSET ESCs expressed comparable levels of mesendoderm and neuro-ectoderm markers while pluripotency markers were barely detectable (Supplementary Fig. 4d). In contrast, expression of shWhsc1.4, which targets the 5' end of Whsc1 mRNA, in ΔSET ESCs caused a delay in pluripotency exit, and EBs derived from these cells also showed the retention of pluripotency gene

expression and the selective reduction of mesendoderm marker expression (Supplementary Fig. 4c, d). As the *Whsc1* gene produces two major transcripts encoding a protein of 1366 amino acids and a short isoform of 648 amino acids that corresponds to the N-terminal portion (Fig. 3a), we examined Whsc1 expression by Western blot, using an antibody recognizing an N-terminal epitope of the protein. We found that ΔSET ESCs express also express two proteins: the larger protein corresponds to a C-terminal truncated form of around 130kD and the short isoform of around 100kD (Supplementary Fig. 4a-b). These results raise the possibility that the N-terminus of Whsc1 plays a role in both the induction of pluripotency exit and mesendoderm specification.

To explore this hypothesis, we generated a complete *Whsc1* knockout ESC line using CRISPR-Cas9 approach. We designed two guide RNAs (gRNAs) that target the exon1 and the exon 15 of *Whsc1* gene leading to the deletion of a ca. 40Kb fragment encoding both the long and short isoforms of *Whsc1* (Fig. 3a). This large deletion was necessary as the *Whsc1* mRNA has multiple AUGs, i.e., potential start codons. As expected, the two homozygous ESC clones (hereafter refer to as *Whsc1*-/-) obtained lack the expression of both isoforms (Fig. 3b). Clone #1 ESCs were used for all further experiments, showing very similar properties to the Whsc1 knockdown cells described before. Thus, *Whsc1*-/- ESCs expressed unaltered levels of Sox2, Oct4 and Nanog (Fig. 3b) and when induced to differentiate in monolayer cultures they showed a delayed down-regulation of *Rex1*, *Oct4* and *Nanog* (Fig. 3c). When used to generate EBs they were found to be impaired in the up-regulation of the mesendoderm but not of the neuroectodermal genes tested (Fig. 3d). Importantly, *Whsc1*-/- EBs exhibited a delayed down-regulation of *Rex1, Pou5f1* and *Nanog* (Fig. 3d).

We next rescued Whsc1 expression with diverse human WHSC1 constructs in our *Whsc1* -/- ESCs. For this, we engineered Flag tagged constructs of full length WHSC1 (WT), two N-terminal constructs of different lengths (Nter-1 and 2) and a C-terminal construct containing the SET-domain (Cter, Fig. 4a). As expected, *Whsc1*-/- cells expressing full length WHSC1 showed rapid down-regulation of *Rex1, Pou5f1* and *Nanog* after induction of differentiation (Fig. 4b). A similar effect was obtained with the two N-terminal constructs but not with the C-terminal construct (Fig. 4b). In line with these findings, expression of full length WHSC1 and the two N-terminal constructs restored mesendoderm up-regulation and pluripotency gene down-regulation during EB formation while the C-terminal construct was ineffective (Fig. 4c).

Our findings show that the ability of Whsc1 to induce efficient exit from pluripotency and mesendoderm specification requires the N-terminus of the protein

and is independent of the protein's methyltransferase function encoded by the C-terminal SET-domain.

Whsc1 activates enhancers of mesendoderm instructive transcription factors together with Brd4

Our observations raise the possibility that Whsc1 directly controls the expression of transcription factors that specify mesodermal and endodermal lineages. We focused on genomic loci encoding the mesendoderm-instructive TFs *Gata4*, *Bra*, *Gata6*, and *Foxa2* and also examined those of the neuroectodermal TFs *Pax6* and *Sox1*. To identify regulatory regions of these genes, we mapped the distribution of the enhancer mark H3K27ac by ChIP-seq in cells derived from Flk1+ mesodermal precursors, hepatocytes and neural stem cells (NSCs). These analyses revealed several putative enhancers selectively marked by H3K27ac in cells from either the mesendodermal or the neuroectodermal lineages (Fig. 5a and Supplementary Fig. 5a).

To explore whether Whsc1 binds to these regions, we performed ChIP analyses using day 6 (D6) EBs. This showed that Whsc1 is significantly enriched at the -16kb region of *Gata4*, the -40kb and -25kb regions of *Bra*, the -106kb and +110kb regions of *Gata6* as well as the -102kb region of *Foxa2* and that this enrichment is reduced in Whsc1-depleted EBs (Fig. 5a, Supplementary Fig. 5a, b). Notably, no significant Whsc1 binding was observed in the regions specifically marked by H3K27ac in the *Pax6* and *Sox1* loci in ectodermal cells (Fig. 5a and Supplementary Fig. 5a). We next cloned several of Whsc1-bound putative enhancers into luciferase reporter constructs, transfected them into either shRNA control or shWhsc1 ESCs and generated EBs with these cells. While all the -16kb *Gata4*, -106kb *Gata6*, -102kb *Foxa2* and -160kb *Pax6* regions were active in D6 control EBs, only those of *Gata4*, *Gata6* and *Foxa2* but not *Pax6* showed decreased luciferase activities in shWhsc1 EBs (Fig. 5b). Importantly, knocking down Whsc1 in D6 EBs resulted in decreased H3K27ac levels in Whsc1-bound regulatory regions identified for *Gata4, Bra, Gata6* and *Foxa2*, whereas no effect was observed for the -160kb *Pax6* and +33kb *Sox1* regions (Fig. 5c). Similar effects were also observed for H3K4me2, another chromatin mark associated with active enhancers (Supplementary Fig. 5b).

The bromodomain protein Brd4 has been reported to interact with Whsc1 [31] and to be required for enhancer activation and gene expression [32]. We therefore performed ChIP experiments to ask whether Brd4 co-occupies the putative enhancers of mesendoderm specific TFs bound by Whsc1. We found that in day 6 EBs Brd4

indeed binds to the -16kb *Gata4*, -40kb *Bra*, -25kb *Bra*, -106kb *Gata6*, 110kb *Gata6* and -102kb *Foxa2* regions but not to the 33kb *Sox1* and -160kb *Pax6* regions (Supplementary Fig. 6a). Importantly, knocking down Whsc1 in EBs resulted in a significant decrease of Brd4 binding to these sites (Supplementary Fig. 6a) and treatment with the Brd4 inhibitor JQ1 [33] showed decreased expression of the mesendodermal regulator genes *Gata4*, *Bra*, *Gata6* and *Foxa2* (Supplementary Fig. 6b). In line with the reported induction of neural differentiation after Brd4 depletion in ESCs [34], we observed an increase of neuroectodermal markers in JQ1 treated EBs (Supplementary Fig. 6c). Moreover, both isoforms of Whsc1 were found to co-immuno-precipitate with Brd4 in D6 EB extracts (Supplementary Fig. 6d), highlighting that the N-terminal part of the Whsc1 is sufficient for this interaction. These results suggest that Whsc1 is required for the binding of Brd4 to Whsc1-targeted mesendoderm enhancers and expression of the corresponding genes.

Taken together, Whsc1 binds to and recruits Brd4 to the putative enhancers of several mesendodermal TFs and controls their activation during differentiation.

Depletion of mesendoderm TFs in ESCs causes a delay in pluripotency exit

The finding that Whsc1 is required for the activation of *Gata4*, *Bra, Gata6*, and *Foxa2* during differentiation raised the possibility that these factors mediate the functions of Whsc1 in pluripotency exit and germ layer specification. To test this hypothesis, we transfected Rex1GFPd2 ESCs with siRNAs against *Gata6*, *Foxa2*, *Gata4* and *Bra,* respectively, resulting in their knockdown (Supplementary Fig. 7a) and induced the cells' differentiation by transfer into N2B27 medium containing Activin A and FBS. As expected, cells transfected with control siRNA showed a complete loss of GFP expression at 48 hours after induction. In contrast, a significant proportion of cells transfected with siRNAs against *Gata6, Bra*, *Gata4 or Foxa2* remained GFP positive (Supplementary Fig. 7b).

To further study the role of the mesendoderm instructive TFs in the exit from pluripotency, we tested *Gata6* and *Foxa2* knockout ESCs generated by CRISPR-Cas9-mediated genome editing. Monitoring gene expression at different times after induction of differentiation revealed that, as *Whsc1-/-* cells, *Gata6-/-* and *Foxa2-/-* cells show a delayed down-regulation of pluripotency-associated genes *Oct4*, *Rex1* and *Nanog* (Fig. 6a). The retention of Oct4 and Nanog expression in *Gata6* and *Foxa2* knockout cells was also confirmed at the protein level in cells induced to differentiate for 72 hours (Fig. 6b). Moreover, when these cells were induced to generate EBs, they

also showed a delayed down-regulation of pluripotency markers as found in *Whsc1-/-* EBs (Fig. 6c).

In conclusion, our results obtained with ESCs depleted or devoid of *Gata6, Bra*, *Gata4 and Foxa2* suggest that these mesendoderm regulators downstream of Whsc1 are required for pluripotency exit and can explain the observed effects of Whsc1 depletion.

Discussion

Here we describe an unexpected new function of the chromatin-related factor Whsc1 in both pluripotency exit and germ layer specification of embryonic stem cells. Depletion of Whsc1 entraps the cells in a pluripotent state and inhibits their specification towards mesoderm and endoderm lineages, without impairing neuro-ectoderm formation. Mechanistically, Whsc1 binds to enhancers of mesendoderm lineage instructive TFs together with Brd4 and is required for the activation of the factors during lineage specification. These functions are independent of the methyltransferase activity of Whsc1 encoded by the SET-domain. Moreover, depletion or ablation of downstream mesendodermal regulators also delayed pluripotency exit, providing an explanation for the effects observed with ESCs lacking Whsc1.

Screens performed in other laboratories have identified several factors required for pluripotency exit, all of which were shown to dismantle the pluripotency-associated regulatory network. Thus, Folliculin together with its interaction partners Fnip1/2 sequesters the pluripotency associated TF Tfe3 in the cytoplasm [8]; the TFs Tcf3 and Zfpn706 repress pluripotency gene expression [9-11]; and the RNA binding proteins Mettl3 and Pum1 destabilize pluripotency gene transcripts [11, 35, 36]. In contrast, we have found that Whsc1 does not act primarily on the pluripotency network, but controls the up-regulation of mesendodermal regulators, including Gata4, Brachyury, Gata6 and Foxa2, which on their own also control pluripotency exit. The novel concept is that instead of becoming ectodermal or die, the induced cells remain stuck in a pluripotent state. Together, our observations suggest that the regulatory networks controlling pluripotency and germ layer specification are intimately linked, challenging the view that pluripotency genes must be down-regulated before lineage specification can ensue [4]. In line with this idea, co-operative binding of pluripotency and somatic cell specific factors has recently been described to be an important intermediate step in reprogramming of fibroblasts to iPSCs [37]. Moreover, it has been shown that pluripotency TFs are crucial for the up-regulation of germ layer genes upon

differentiation, such that differentiation fails when pluripotency TFs are acutely ablated [7, 38-40]

Future studies are required to address the question how Whsc1 is involved in the activation of enhancers. Our experiments so far have shown that it interacts with the enhancer-associated Brd4 protein and that Whsc1 is required to maintain the binding of Brd4 to enhancers of mesendoderm regulators. However, it remains unclear how Whsc1 becomes recruited to mesendoderm enhancers in the first place and whether it remains bound in cells treated with the Brd4 inhibitor JQ1.

In certain forms of cancer, including multiple myeloma and acute lymphoblastic leukaemia, WHSC1 has been found to be overexpressed or hyper-activated, resulting in an increased methylation of H3K36 on promoters of oncogenes that drive the disease [28, 29, 41]. This is mediated by the methyltransferase activity encoded in the SET-domain. However, NSD3, another member of the Nsd family, has been found to act via SET-independent mechanisms. Thus, in a subset of acute myeloid leukaemia a short form of NSD3 recruits Brd4 to enhancers of genes that drive the disease[42]. Interestingly, the PWWP domain responsible for this interaction is highly conserved between members of the Nsd family, including Whsc1 where it is located at the N-terminus that interacts with Brd4.

Wolf-Hirschhorn syndrome patients with *WHSC1* deletions often exhibit symptoms characteristic of midline defects, including craniofacial malformations and heart defects [43]. These features have been attributed to the SET-domain of the protein, since *Whsc1-/-* mice lacking the SET-domain largely reflect this phenotype, exhibiting aberrant craniofacial, cardiac and cartilage structures [30, 44]. Our observation that ESCs lacking the SET-domain are moderately impaired in the formation of cardiac cells, while those depleted for the entire protein are strongly inhibited, support the idea that both SET-dependent and independent functions contribute to cardiac differentiation. Future work will be necessary to investigate the SET-domain independent role of Whsc1 during germ layer specification and cancer.

Methods

Cell culture

Mouse ESCs were cultured on plastic dishes pre-coated with 0.1% (W/V) EmbryoMax® gelatine (EMD Millipore) with either DMEM/F12 medium and Neurobasal® medium mixed at 1:1 ratio, supplemented with MEM non-essential amino acids solution (1X),

sodium pyruvate (1mM), L-glutamine (2mM), Penicillin (100 U/mL), Streptomycin (100µg/mL), beta-mercaptoethanol (ME, 50µM), N2 and B27 supplements, the small-molecule inhibitors PD0325901 (1µM, Selleck Chemicals) and CHIR99021 (3 µM, Selleck Chemicals), and ESGRO® Leukaemia Inhibitory Factor (LIF) (1000 U/mL, EMD Millipore), or with Knockout™ DMEM medium supplemented with 15% (V/V) ES cell qualified foetal bovine serum (ES FBS), MEM non-essential amino acids (1X), sodium pyruvate (1mM), L-glutamine (2mM), Penicillin (100 U/mL), Streptomycin (100µg/mL) and 50µM ME. All culture reagents were from Life Technologies unless noted otherwise.

RNA interference assay

The pLKO-TRC based shRNA constructs against Whsc1, Cbx4, L3mbtl3 and the scrambled shRNA were purchased from Sigma (Supplementary Table 2). Lentiviral particles were produced by co-transfecting the constructs with pCMVΔ8.9 and pVSVG vectors in 293T cells. Forty-eight hours later the supernatant was collected and filtered through a 0.22µm filter, concentrated by ultra-centrifugation with 22,000 rcf at 22°C for 2 hours. Precipitated lentiviral particles were suspended in phosphate-buffered saline solution (PBS) on a shaker at 4°C for 1 hour and stocked at -80°C. For lentivirus infection, mESCs were disassociated using Accutase® (Life Technology), washed twice with PBS, then $0.5X10^6$ cells were suspended with 100µL medium without inhibitors or LIF, together with concentrated virus and 4µg/mL polybrene (Sigma) and centrifuged with 300g at 37°C for 90 min. Infected cells were then transferred to gelatine pre-coated dishes filled with complete medium including inhibitors and LIF. Three days after, infected cells were selected using puromycin (2µg/mL) for additional 3 days. Twenty pmoles of ON-TARGETplus SMART pool siRNAs against *Gata6*, *Foxa2*, *Gata4*, *Brachyury* and control siRNA (GE Dharmacon) (Supplementary Table 2) were electroporated into ESCs using Nucleofector™ according manufacturer's instructions.

Genome editing with CRISPR-Cas9

The two gRNAs targeting the Whsc1 locus were designed using CRISPR design tool from MIT (http://crispr.mit.edu) (Supplementary Table 2). After annealing the two gRNAs were cloned into the DECKO (Double Excision CRISPR Knockout) version of SpCas9(BB)-2A-GFP (PX458) (original backbone from Addgene #48138) [45]. The DECKO system facilitates genomic deletion by delivering simultaneously two gRNAs plus Cas9 into cells [46]. Twenty-four hours after transfection, single cells were sorted

by FACS. Mutant clones were screened through PCR and sequencing to identify homozygously targeted clones.

To generate *Gata6* and *Foxa2* KO ESCs, gRNAs were cloned into the p330 plasmid (https://www.addgene.org/42230/). Together with the gRNA cells were tagged using a constitutive GFP. Both gRNAs resulted in large deletion and subsequent frame shift and details could be found in (Goldmann et al. in preparation).

Alkaline phosphatase assay

Alkaline phosphatase assays were performed using kit from Stemgent, according to the manufacturer's instruction

Assay for pluripotency exit

ESCs were disassociated with Accutase for 2 min at room temperature, washed twice with PBS, and $2X10^5$ cells were seeded in 6-well plates pre-coated with 0.1% (W/V) EmbryoMax® gelatine in culture medium containing N2B27 supplements plus Activin A (25ng/mL, Peprotech) and 10% FBS without the two inhibitors or LIF. GFP expression level was monitored by flow cytometry on a LSR Fortessa analyser (BD Bioscience) and the analysis was performed using Diva v6.1.2 (BD Biosciences) and FlowJo software v10.0.6 (TreeStar). In some experiments Rex1GFPd2 cells that retain GFP expression were selected by adding 10µg/mL blasticidin (Sigma) to the medium.

Embryoid body formation, cardiac progenitor, definitive endoderm and neural progenitor differentiation

ESCs were induced to form EBs with "EB differentiation medium" (Knockout™ DMEM medium supplemented with 10% ES FBS, MEM non-essential amino acids (1X), sodium pyruvate (1mM), L-glutamine (2mM), Penicillin (100 U/mL), Streptomycin (100µg/mL) and 50µM ME and plated as hanging drops (400 cells in 20µL) on the lid of a 100 mm Petri dish. The cultures were harvested 3 days later and transferred into bacterial dishes for further culture in suspension. The cardiac progenitor differentiation assay was performed as described [15] with slight modifications. Briefly, ESCs were seeded in hanging drop EB cultures with DMEM medium containing 10% serum to induce EB formation. Three days later, at least 100 EBs were transferred individually into separate wells of a gelatine pre-coated 96-well plate supplemented with 100 µL "EB differentiation medium" and cultured for another 3 days before counting beating EBs under low power magnification. The definitive endoderm and neural differentiation protocols were performed as described [26] and [16].

Teratoma formation

Teratoma formation assays were performed as described previously [47]. Briefly, knockdown and control ESCs grown in 2i plus LIF conditions ($0.5X10^6$) were injected subcutaneously into SCID mice. Four weeks later, mice with tumours were euthanized and tumours were either snap-frozen in liquid nitrogen for further gene expression analysis or fixed in formalin, embedded in paraffin, sectioned and stained with hematoxylin and eosin for histological analysis. All experimental procedures were approved by the local ethical committee (CEEA - PRBB), and met the guidelines of the local (Catalan law 5/1995 and Decrees 214/97, 32/2007) the European regulations (EU directives 86/609 and 2001–486) and the Standards for Use of Laboratory Animals A5388-01 (NIH).

Chromatin Immunoprecipitation Assay

ChIP assays in EBs were performed using True MicroChIP kit (Diagenode) according to manufacturer's instruction. EBs were cross linked with 4% formaldehyde for 20 min for H3K27Ac and H3K4me2 or "double-cross linked" with DSG 2mM for 30 min then with 4% formaldehyde for 20 min for Brd4 and Whsc1. Genomic DNA was purified using MicroChIP DiaPure columns. Antibodies and primers used are listed in Supplementary Table 2. ChIP-seq data of H3K27ac in Flk1+ mesodermal progenitors and in hepatocytes were from [48] (GSM1359831) and [49] (SRX185861) respectively. ChIP-seq of H3K27ac in neural stem cells was performed as previously described [50] and analysed accordingly [51]; the dataset can be found at GSM1463937.

Luciferase Assay

The enhancers of *Foxa2*, *Gata4*, *Gata6* and *Pax6* were amplified by PCR, digested using XhoI and PstI restriction enzymes and cloned into the CSI–LUC2–minP vector. ESCs were transfected with a β-galactosidase control plasmid and the respective enhancer plasmids or the empty vector as control using TransIT transfection reagent (Mirus). Luciferase and β-galactosidase activities were measured using the luciferase assay system (Promega) and the β-galactosidase enzyme assay system (Promega) according to the manufacturer's protocol. All experiments were carried out in triplicate and normalized to β-galactosidase activity.

Western Blot and Co-immunoprecipation Analysis

Proteins were extracted using NET-2 buffer (50mM Tris pH 7.4, 200mM NaCl, 0,1% triton and proteases inhibitors) and loaded on SDS-PAGE gel. For endogenous Co-IP experiments, lysates were prepared using NET-2 buffer and centrifuged to eliminate

insoluble material. The extract was pre-cleared by incubation with magnetic beads for 2h on a rotating wheel at 4°C. To crosslink the antibodies to the Dynabeads® A or G (Life Technology) magnetic beads, 50µl of beads were incubated with 3ug of antibody or IgGs during 2 hrs at room temperature on a wheel. The mix was then washed once with PBS and twice with triethanolamine (pH8,2, 0,2M) and then incubated during 20min at room temperature with dimethyl pimelimidate (DMP, 20m). To stop the reaction two 5 min washes with Tris-Cl (pH8, 50mM) were performed followed by 3 washes with PBS. Then beads were incubated with citric acid (0,1M, pH3) for 2 min. Finally, the mix was washed twice with lysis buffer before incubating it with the precleared protein extracts overnight at 4°C on a wheel. Afterwards 1/20th of the mixture was kept as the unbound fraction and the rest washed six times with lysis buffer. To elute, the last wash was removed and the beads resuspended in 1X Laemmli buffer without ME and heated for 20 min at 60°C. Magnetic beads were then separated, supernatant taken, complemented with 5% ME and boiled before loading on an SDS-PAGE acrylamide gel to analyse bound proteins by Western Blotting.

Gene expression analysis

Total RNA was isolated using RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. Double strand complementary DNA (cDNA) was synthesized using the High Capacity RNA-to-cDNA kit (Applied Biosystems). Real-time Polymerase Chain Reaction (PCR) was performed on ViiA™ 7 system using Power SYBR® Green PCR Master Mix (Applied Biosystems). The relative gene-expression levels were calculated by the $2^{-\Delta\Delta Ct}$ method. Primers can be found in the Supplementary Table 2.

Statistics

Statistical analysis was performed by the two-tailed unpaired Student's t test. All values presented were expressed as mean ± SD. *=P<0.05, **=P<0.01 and ***=P<0.001.

Figure Legends

Figure 1. An *in silico* screen reveals a role of Whsc1 in the exit from pluripotency of ESCs.

(a) Experimental procedure to assess exit from pluripotency using Rex1GFPd2 ESCs. (**b**) Representative FACS plots of GFP (Rex1) expression in cells expressing scrambled shRNA (shScr) as a control or shWhsc1 knockdown constructs (shWhsc1.2 and shWhsc1.4) at 0 hr and 72 hrs after induction (n=3). (**c**) Representative images of ESCs expressing shScr and shWhsc1.4 48 hours after induction of exit from

pluripotency. Scale bar: 500 μm (**d**) Expression of Whsc1 quantified by qRT-PCR (left panel) and Western blot (right panel) in control (shScr) and Whsc1 knockdown ESCs. Beta Tubulin was used as loading control. (**e**) Expression of pluripotency genes 72 hrs after induction of exit from pluripotency (n=3).

Figure 2. Whsc1 is required for mesendoderm differentiation.

(**a**) Schematics of EB differentiation assay. (**b**) Images of Day 6 EBs derived from control and Whsc1 depleted Rex1GFPd2 cells (left panel) and quantification of EB size. Average and SD are indicated. Scale Bar: 500 μm. (**c**) Expression of Whsc1, pluripotency, mesendoderm, ectoderm markers in Day 6 EBs quantified by RT-qPCR from control and Whsc1 depleted Rex1GFPd2 cells (n=3). (**d**) Expression of Whsc1 and pluripotency factors in Day 6 EBs evaluated by Western Blot from control and Whsc1-depleted Rex1GFPd2 cells. Beta-Tubulin was used as loading control. (**e**) Schematics of EB replating assay and images of alkaline phosphatase (AP) stained colonies obtained from control and Whsc1-depleted EBs. Scale bar: 1mm; Right panel: quantification of AP positive colonies (n=2).

Figure 3. Ablation of *Whsc1* results in delayed down-regulation of pluripotency genes and impairment of mesendoderm formation

(**a**) Schematics of strategy for generating ESCs with a complete *Whsc1* knockout using CRISPR-Cas9 approach. Top panel: organization of *Whsc1* that can generate two transcript isoforms (long and short). The targets of gRNAs (exon1 and exon15) are indicated; Middle panel: genomic sequences of wild-type *Whsc1* (*Whsc1* +/+) spanning exon1 and exon 15 (exon1 and exon15 are indicated as red boxes); Bottom panel: sequence and chromatograph of the Whsc1 knockout (*Whsc1*-/-). (**b**) Western blot of Whsc1 and pluripotency factors, Sox2, Oct4 and Nanog in *Whsc1* +/+ and two *Whsc1* -/- ESC clones. Beta tubulin was used as loading control. (**c**) Expression kinetics of pluripotency genes in *Whsc1* +/+ and *Whsc1* -/- cells were monitored by RT-qPCR after transfer into N2B27 medium containing Activin A and FBS (n=3). (**d**) Expression kinetics of mesendodermal, ectodermal and pluripotency genes were evaluated by RT-qPCR during EB formation using *Whsc1* +/+ or *Whsc1* -/- ESCs (n=3).

Figure 4. N-terminus of Whsc1 is sufficient to rescue pluripotency exit and mesendoderm differentiation in *Whsc1* -/- cells.

(**a**) Schematics of full length and truncated human WHSC1 constructs tagged by 3XFlag. (**b**) Expression kinetics (RT-qPCR) of pluripotency genes in *Whsc1* +/+ and in *Whsc1* -/- cells infected with WT, Nter-1, Nter-2 or Cter WHSC1 constructs after transfer into N2B27 medium containing Activin A and FBS (n=3). (**c**) Expression of

mesendodermal, ectodermal and pluripotency genes were by RT-qPCR in Day 6 EBs derived from *Whsc1* +/+ and *Whsc1* -/- ESCs infected with WT, Nter-1, Nter-2 or Cter WHSC1 constructs (n=3).

Figure 5. Whsc1 controls the enhancer activity of mesendoderm transcription factors.

(**a**) Left panel: H3K27ac ChIP-seq profiles on the loci of *Brachyury, Foxa2 and Pax6* in Flk1+ mesoderm progenitors [48], hepatocytes [49] and NSCs. The lines below each panel correspond to putative enhancers up or downstream the respective gene, with numbers indicating distance to the TSS in kb. Whsc1-bound putative enhancers are indicated by bleu boxes; Right panel: ChIP-qPCR quantification of Whsc1 occupancy on the same enhancers as shown on the left compared to IgG in Day 6 EBs (n=2). (**b**) Luciferase assay of the putative regulatory regions of *Gata4, Gata6, Foxa2 and Pax6* in Day 6 control or Whsc1-depleted EBs (n=2). (**c**) H3K27ac enrichments on putative enhancers of lineage-instructive TFs were quantified by ChIP-qPCR in Day 6 control or Whsc1-depleted EBs (n=2).

Figure 6. Mesendoderm transcription factors are required for efficient pluripotency exit.

(**a**) Expression kinetics of pluripotency genes in *Whsc1* -/-, *Gata6* KO, *Foxa2* KO and control ESCs (WT) monitored by RT-qPCR after transfer into N2B27 medium containing Activin A and FBS (n=3). (**b**) Western blot analysis of Oct4 and Nanog in cells described in (a). Beta tubulin was used as loading control. (**c**) Expression kinetics of pluripotency genes evaluated by RT-qPCR during EB formation using *Whsc1* -/-, *Gata6* KO and *Foxa2* KO and control ESCs (WT) (n=3).

Supplementary figure legends

Supplementary Figure 1. Whsc1 is involved in the exit from pluripotency of mESCs.

(**a**) Up-regulation of Chromatin-related Factors (CRFs) during embryoid body (EB) differentiation of R1 (GSE2972) (left panel) and J1 ESCs (GSE3749) (right panel) [52]. Selected candidate genes, *Cbx4*, *L3mbtl3* and *Whsc1,* are indicated. (**b**) Expression of pluripotency, mesendoderm and neuroectoderm markers during exit from pluripotency induced by N2B27, Activin A and FBS. Expression fold changes relative to 0 hr are indicated as bold numbers. (**c**) Left panel: Evaluation of knockdown efficiency by RT-qPCR after shRNA transductions in Rex1GFPd2 cells (n=2). Right

panel: Flow cytometry analysis of GFP (Rex1) expression in Rex1GFPd2 cells depleted of *Cbx4* or *L3mbtl3* at 0 hr and 72 hrs after induction. (**d**) Evaluation of serial induction of exit from pluripotency. Left-top panel: schematic representation of experimental procedure; left-bottom panel: quantification of AP+ colonies (n=2); Right panel: representative images of control (shScr) and Whsc1 depleted Rex1GFPd2 cells subjected to two rounds of exit from pluripotency. Differentiated cells (cells lost GFP/Rex1 expression) were eliminated by Blasticidin and pluripotent cells were stained by AP. Sacle bar: 500 μm (**e**) Cell proliferation assay of control (shScr) and Whsc1 depleted (shWhsc1.2 and shWhsc1.4) Rex1GFPd2 cells (n=2). (**f**) Evaluation of pluripotency gene expression in control (shScr) and Whsc1 depleted (shWhsc1.2 and shWhsc1.4) Rex1GFPd2 cells by RT-qPCR (n=3). (**g**) Western blot analysis of pluripotency factors, Nanog, Sox2 and Oct4 in control (shScr) and Whsc1-depleted (shWhsc1.2 and shWhsc1.4) Rex1GFPd2 cells. Beta tubulin was used as loading control.

Supplementary Figure 2. Whsc1 depletion results in reduced mesendoderm differentiation in teratomas *in vivo*.

(**a**) HE staining of teratomas from control (shScr) and Whsc1 depleted (Whsc1.2 and Whsc1.4) cells. Ten teratomas from each group were examined, and representative images are shown. Endodermal (gut), mesodermal (cartilage) and ectodermal (neuroepithelia) were indicated on the figures. Scale Bar: 200 μm. (**b**) Expression of *Whsc1*, mesendoderm, ectoderm and pluripotency makers was quantified by RT-qPCR in control (shScr) and Whsc1 depleted teratomas (shWhsc1.2 and shWhsc1.4) (n=3).

Supplementary Figure 3. Whsc1 is required for efficient generation of cardiac progenitors and definitive endoderm but not of neural progenitor cells.

(**a**) Percentage of EBs with beating cells from control and Whsc1-depleted ESCs counted at Day 6 to 10 after induction of differentiation (n=2). (**b-c**) RT-qPCR quantification of the expression of cardiac progenitor and pluripotency markers 10 days after induction of differentiation (n=2). (**d**) Definitive endoderm differentiation using Eomes-GFP ESCs. FACS profiles of ESCs 0 and 7 days after induction of differentiation. (**e**) Quantification of Eomes-GFP positive cells in control and Whsc1-depleted cells 7 days after induction of differentiation (n=3). (**f-g**) Expression levels by RT-qPCR of definitive endoderm (f) and pluripotency markers (g) during induction of definitive endoderm (n=3) in shScr and shWhsc1.4 cells. (**h**) Expression levels by RT-qPCR of neural markers at several time points after neural progenitor induction (n=3)

and Quantification of pluripotency gene expression by RT-qPCR 8 days after induction of differentiation (n=3).

Supplementary Figure 4. The SET-domain is dispensable for the functions of Whsc1 in pluripotency exit and mesendodermal differentiation.

(**a**) Schematics of wild type (WT) and the SET-domain deleted (ΔSET) Whsc1 proteins (ΔSET) [30]. (**b)** Western blots of Whsc1 expression in WT and ΔSET cells, as well as ΔSET cells with control or Whsc1 shRNA (shWhsc1.4). (**c**) Expression levels by RT-qPCR of Whsc1 and pluripotency genes in WT and ΔSET cells, as well as in ΔSET cells expressing shScr or shWhsc1.4 constructs at 48 and 72hrs after induction of exit from pluripotency by transferring cells into N2B27 medium containing Activin A and FBS (n=3). (**d**) Expression levels by RT-qPCR of Whsc1, mesendoderm, ectoderm markers and pluripotency markers in Day 6 EBs derived from WT and ΔSET ESCs, as well as in ΔSET cells with shScr or shWhsc1.4 constructs (n=3).

Supplementary Figure 5. Whsc1 binds specifically to the enhancers of mesendoderm TFs.

(**a**) Left panel: H3K27ac ChIP-seq profiles on the loci of *Gata4, Gata6 and Sox1* in Flk1+ mesoderm progenitors [48] (GSM1359831), hepatocytes [49] (SRX185861) and NSCs. The lines below each panel correspond to putative enhancers up or downstream the respective gene, with numbers indicating distance to the TSS in kb. The Whsc1-bound putative regulatory regions are marked by blue boxes. Right panel: ChIP-qPCR quantification of Whsc1 occupancy on the same enhancers as shown on the left compared to IgG in Day 6 EBs (n=2). (**b**) ChIP-qPCR quantification of IgG (top panel), H3K4me2 (middle panel) and Whsc1 (bottom panel) occupancies on putative regulatory regions of *Gata4, Brachyury, Gata6, Foxa2, Pax6 and Sox1* in D6 control (shScr) and Whsc1-depleted (shWhsc1.4) EBs (n=2).

Supplementary Figure 6. Whsc1 interacts with Brd4 on enhancers of mesendoderm transcription factors.

(**a**) ChIP-qPCR quantification of Brd4 occupancy on Whsc1-bound putative enhancers of *Gata4, Bra, Gata6, Foxa2* and control regions (*Sox1* and *Pax6*) in Day 6 shScr and shWhsc1.4 EBs (n=2). (**b-c**) Expression of mesendodermal markers (b) and ectodermal markers (c) quantified by RT-qPCR in Day 3 control or JQ1-treated EBs (n=3). (**d**) Western blots of endogenous Whsc1 (~180 and ~100kDa isoforms) and Brd4 co-immunoprecipitated from total protein extracts of Day 6 EBs. Inputs correspond to 10% of total extract.

Supplementary Figure 7. Mesendoderm TFs are required for efficient exit from pluripotency.

(**a**) siRNA knockdown efficiency was evaluated by RT-qPCR in Rex1GFPd2 cells 72 hrs after induction of exit from pluripotency (n=2). (**b**) Kinetics of GFP expression determined by FACS in Rex1GFPd2 ESCs transfected with siRNAs against *Bra*, *Gata6*, *Gata4*, *Foxa2* at different times after differentiation induction (n=2).

Author Contributions

T.V.T and T.G. conceived the project, designed the experimental work and wrote the manuscript. T.V.T, B.D.S., A.D., J.L.S., C.S.M. L.D.A.A. and J.G. performed the experiments. A.G.M conducted the bioinformatics analysis.

Acknowledgements

We would like to thank A. Smith for providing the Rex1GFPd2 cells, K. Nimura for the Whsc1 ΔSET cells, E. J. Robertson for the Eomes-GFP cells, J. E. Bradner for the JQ1 compound and R. A. Young for ChIPseq data; L. Batlle, C. Berenguer for technical help; the CRG core facilities flow cytometry, advanced light microscopy and tissue engineering; L. Di Croce, B. Payer, M. Behringer, J. Licht, K. M. Loh, K. Kaji, G. Stik and R. Stadhouders for advice and discussions. T.V.T. and J.L.S. were supported by Juan de la Cierva postdoctoral fellowships (MINECO, FJCI-2014-22946 and IJCI-2014-21872), B.D.S. by an EMBO long-term fellowship (#ALTF 1143-2015), J.G. by a Boehringer Ingelheim Graduate Student Fellowship. This work was supported by the EU-FP7 project BLUEPRINT, the Spanish Ministry of Economy and Competitiveness, Centro de Excelencia Severo Ochoa 2013-2017 and the CERCA Program Generalitat de Catalunya. T.V.T, J.L.S. and B.D.S were supported by a CRG award for junior collaborative projects.

References

1. Nichols, J. & Smith, A. The origin and identity of embryonic stem cells. *Development* **138**, 3-8 (2011).
2. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567-582 (2008).
3. Young, R.A. Control of the embryonic stem cell state. *Cell* **144**, 940-954 (2011).

4. Martello, G. & Smith, A. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol* **30**, 647-675 (2014).
5. Clevers, H., Loh, K.M. & Nusse, R. Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science* **346**, 1248012 (2014).
6. Loh, K.M., Lim, B. & Ang, L.T. Ex uno plures: molecular designs for embryonic pluripotency. *Physiol Rev* **95**, 245-295 (2015).
7. Loh, K.M. & Lim, B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* **8**, 363-369 (2011).
8. Betschinger, J. *et al.* Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell* **153**, 335-347 (2013).
9. Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol* **13**, 838-845 (2011).
10. Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11**, 491-504 (2012).
11. Leeb, M., Dietmann, S., Paramor, M., Niwa, H. & Smith, A. Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell* **14**, 385-393 (2014).
12. Ho, L. & Crabtree, G.R. Chromatin remodelling during development. *Nature* **463**, 474-484 (2010).
13. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-1155 (2013).
14. Zhang, Y. *et al.* H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm. *Cell Rep* **8**, 1989-2002 (2014).
15. Morey, L. *et al.* Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms. *Cell Stem Cell* **17**, 300-315 (2015).
16. Aloia, L. *et al.* Zrf1 is required to establish and maintain neural progenitor identity. *Genes Dev* **28**, 182-197 (2014).
17. Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47-62 (2012).
18. Arai, S. & Miyazaki, T. Impaired maturation of myeloid progenitors in mice lacking novel Polycomb group protein MBT-1. *EMBO J* **24**, 1863-1873 (2005).
19. Bergemann, A.D., Cole, F. & Hirschhorn, K. The etiology of Wolf-Hirschhorn syndrome. *Trends Genet* **21**, 188-195 (2005).
20. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**, 183-186 (2003).
21. Fei, T. *et al.* Smad2 mediates Activin/Nodal signaling in mesendoderm differentiation of mouse embryonic stem cells. *Cell Res* **20**, 1306-1318 (2010).
22. Wang, Q. *et al.* The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* **20**, 70-86 (2017).

23. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909-918 (2008).
24. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532 (2011).
25. Costello, I. *et al.* The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nat Cell Biol* **13**, 1084-1091 (2011).
26. Morrison, G., Scognamiglio, R., Trumpp, A. & Smith, A. Convergence of cMyc and beta-catenin on Tcf7l1 enables endoderm specification. *EMBO J* **35**, 356-368 (2016).
27. Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525-2534 (2013).
28. Kuo, A.J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol Cell* **44**, 609-620 (2011).
29. Popovic, R. *et al.* Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. *PLoS Genet* **10**, e1004566 (2014).
30. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* **460**, 287-291 (2009).
31. Sarai, N. *et al.* WHSC1 links transcription elongation to HIRA-mediated histone H3.3 deposition. *EMBO J* **32**, 2392-2406 (2013).
32. Devaiah, B.N. *et al.* BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat Struct Mol Biol* **23**, 540-548 (2016).
33. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073 (2010).
34. Di Micco, R. *et al.* Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep* **9**, 234-247 (2014).
35. Batista, P.J. *et al.* m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707-719 (2014).
36. Geula, S. *et al.* Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002-1006 (2015).
37. C hronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442-459 e420 (2017).
38. Teo, A.K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev* **25**, 238-250 (2011).
39. Loh, K.M. & Lim, B. Epigenetics: Actors in the cell reprogramming drama. *Nature* **488**, 599-600 (2012).
40. Radzisheuskaya, A. *et al.* A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. *Nat Cell Biol* **15**, 579-590 (2013).
41. Jaffe, J.D. *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet* **45**, 1386-1391 (2013).
42. Shen, C. *et al.* NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler. *Mol Cell* (2015).

43. Rutherford, E.L. & Lowery, L.A. Exploring the developmental mechanisms underlying Wolf-Hirschhorn Syndrome: Evidence for defects in neural crest cell migration. *Dev Biol* **420**, 1-10 (2016).
44. Lee, Y.F., Nimura, K., Lo, W.N., Saga, K. & Kaneda, Y. Histone H3 lysine 36 methyltransferase Whsc1 promotes the association of Runx2 and p300 in the activation of bone-related genes. *PLoS One* **9**, e106661 (2014).
45. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013).
46. Aparicio-Prat, E. *et al.* DECKO: Single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. *BMC Genomics* **16**, 846 (2015).
47. Di Stefano, B. *et al.* C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506**, 235-239 (2014).
48. Org, T. *et al.* Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. *EMBO J* **34**, 759-777 (2015).
49. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355-364 (2014).
50. Kagey, M.H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435 (2010).
51. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).
52. Hailesellasse Sene, K. *et al.* Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics* **8**, 85 (2007).