Whsc1 links pluripotency exit with mesendoderm specification

Tian V. Tian, Bruno Di Stefano, Grégoire Stik, Maria Vila-Casadesús, José Luis Sardina, Enrique Vidal, Alessandro Dasti, Carolina Segura-Morales, Luisa De Andrés-Aguayo, Antonio Gómez, Johanna Goldmann, Rudolf Jaenisch and Thomas Graf

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 (also known as Nsd2 and MMSET) has a dual role in pluripotency exit and germ layer specification of embryonic stem cells. On induction of differentiation, a proportion of Whsc1-depleted embryonic stem cells remain entrapped in a pluripotent state and fail to form mesendoderm, although they are still capable of generating neuroectoderm. These functions of Whsc1 are independent of its methyltransferase activity. Whsc1 binds to enhancers of the mesendoderm regulators Gata4, T (Brachyury), Gata6 and Foxa2, together with Brd4, and activates the expression of these genes. Depleting each of these regulators also delays pluripotency exit, suggesting that they mediate the effects observed with Whsc1. Our data indicate that Whsc1 silencing of the pluripotency regulatory network with activation of mesendoderm lineages.

Embryonic stem cells (ESCs) are an excellent model for studying the generation of the three main germ layers, endoderm, mesendoderm and ectoderm. The first event during the transition of ESCs into more differentiated cells is the exit from pluripotency, which includes silencing of the core pluripotency transcription factors Oct4, Sox2 and Nanog. This process can be externally initiated either by deprivation of self-renewal signals or by exposure of the cells to differentiation-inducing cues. Subsequently, germ layer-instructive transcription factors become upregulated and activate markers that define the various lineages. These factors comprise Brachyury (T) and Gata4 for mesoderm, Gata6, Gata4 and Foxa2 for endoderm, and Pax6 and Sox1 for neuroectoderm. Screens with ESC reporter lines and haploid ESCs have identified several regulators required for pluripotency exit, including several transcription factors and RNA-binding proteins, all of which were shown to affect members of the pluripotency network.

Besides transcription factors and non-coding 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. 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. Others act in a more restricted fashion, as exemplified by the requirement of Setd2 for endodermal differentiation, Mel18 for mesodermal differentiation and Zrf1 for neuroectodermal differentiation.

Here, we report that Whsc1 is required for efficient pluripotency exit and mesendoderm specification, independent of its methyltransferase activity. These effects can be explained by its ability to activate several mesendoderm regulators.

Results

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 that record 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 long non-coding RNAs (Supplementary Table 1 and Supplementary Fig. 1a). Three candidates consistently scored among the most upregulated genes. These were Cbx4 (encoding a Polycomb group protein that has been shown to orchestrate ESC differentiation), L3mbtl3 (a putative Polycomb group family member) and Whsc1 (encoding a SET-domain methyltransferase, which in humans is associated with Wolf–Hirschhorn syndrome).

To study whether these candidates are required for pluripotency exit, we tested the effects of knockdowns on the differentiation of a pluripotency reporter ESC line. This line, which contains a destabilized version of green fluorescent protein (GFP) knocked into the Rex1 locus (Rex1GFPd2), was induced to differentiate by transfer into medium containing N2 and B27 supplements, activin A and foetal bovine serum (FBS). On induction, the reporter line became GFP negative within 48–72 h (Fig. 1a–c), confirming that the loss of Rex1 expression is a highly sensitive readout for exit from the pluripotent state. Concomitantly, the cells downregulated pluripotency-associated genes and upregulated 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 candidate genes. After induction of differentiation, cells expressing scrambled shRNA (shScr) became GFP negative within 48–72 h post-induction, as did cells expressing shCbx4 and shL3mbtl3 (Supplementary Fig. 1c). In contrast, 25–35% of ESCs transduced with the shWhsc1 constructs retained GFP expression (Fig. 1b–c). Notably, both Whsc1 knockdown constructs strongly reduced the expression of both Whsc1 messenger RNA (mRNA) and the ~180 and ~100kDa protein isoforms (Fig. 1d). Interestingly, shWhsc1
Depletion of Whsc1 in ESCs resulted in neither an upregulation of pluripotency factors nor an alteration of the cells' growth kinetics (Fig. 2c). In addition, Whsc1-depleted cells, when re-plated after induced differentiation, yielded large numbers of colonies positive for alkaline phosphatase (an indicator of pluripotency), even after two consecutive rounds of treatment, whereas no such colonies could be recovered from control cells as early as after one round (Supplementary Fig. 1d). Depletion of Whsc1 in ESCs resulted in neither an upregulation of pluripotency factors nor an alteration of the cells' growth kinetics (Supplementary Fig. 1e–g), ruling out more indirect effects on pluripotency exit.

In summary, Whsc1 depletion impaired the downregulation of pluripotency markers, indicating a requirement for efficient pluripotency exit.

Depletion of Whsc1 impairs induction of mesendoderm differentiation in vitro and in vivo. To evaluate the effect of Whsc1 depletion on the differentiation capacity of ESCs, we examined EBs derived from knockout ESCs (Fig. 2a). Whsc1-depleted EBs were smaller than those from control ESCs (Fig. 2b), and exhibited lower levels of Whsc1 mRNA and protein while retaining pluripotency factor expression (Fig. 2c,d). Importantly, Whsc1-depleted EBs also showed strongly reduced expression of the mesendoderm regulator genes Gata4, T, Gata6, Foxa2, Sox17 and Flk1, while neuroectodermal gene expression (Sox1, Pax6 and Nes) remained unaffected (Fig. 2c). In addition, when shWhsc1.2 and shWhsc1.4 at 0 and 72 h after induction. Numbers in the plots represent percentage of GFP positive cells. Representative fluorescence-activated cell sorting plots of GFP (Rex1) expression in cells expressing shScr as a control or shWhsc1 knockdown ESCs. Tubulin was used as loading control. Scanned images of unprocessed blots are shown in Supplementary Fig. 8.

To probe their differentiation potential in vivo, we generated teratomas in immunodeficient mice. Control cell-derived teratomas contained tissues from all three germ layers, including muscle, cartilage, gut epithelium and neuro-epithelium. In contrast, derived Whsc1-depleted teratomas were mainly composed of neuro-epithelium and poorly differentiated cells (Supplementary Fig. 2b). Accordingly, shWhsc1 teratomas retained pluripotency and neuroectoderm marker expression while showing reduced levels of mesendoderm markers (Supplementary Fig. 2c).

Since both EBs and teratomas comprise a complex mixture of cells, we next tested the effects of Whsc1 knockdown on the directed differentiation of ESCs towards mesoderm, endoderm and ectoderm. Whsc1-depleted cells generated significantly fewer beating cardiomyocyte-containing colonies than control cells (Supplementary Fig. 3a and Supplementary Videos 1 and 2) and a reduced expression of the cardiac regulators Mef2c and Nkx2.5 at 10 d postinduction, while exhibiting increased Rex1 and Pou5f1 expression (Supplementary Fig. 3b,c). To test the effects of shWhsc1 on definitive endoderm differentiation, Eomes-GFP reporter ESCs were transferred into medium containing activin A, Fgf4, heparin, PI-103 and CHIR27. 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 showed impaired upregulation of the endoderm markers Cxcr4, Cldn6, Foxa2, Gata6 and Sox17, and inefficient downregulation of
Pou5f1 and of Esrrb requires the catalytic SET domain on the carboxy (C) terminus to pluripotency exit and mesendoderm differentiation. Whsc1 is dispensable for efficient impairment of mesendoderm differentiation. The catalytic SET domain of Whsc1 is required for efficient pluripotency exit induced by Whsc1 depletion is coupled to an directed germ layer induction experiments confirm that the delay in the expression of the neural markers Pax6 and Rex1, the pluripotency regulators Pou5f1 and Esrrb (Supplementary Fig. 3f,g). Finally, we derived neural progenitor cells from control and Whsc1-depleted Rex1GFPd2 cells (Supplementary Fig. 3h).

Taken together, the results obtained with EBs, teratomas and directed germ layer induction experiments confirm that the delay of pluripotency exit induced by Whsc1 depletion is coupled to an impairment of mesendoderm differentiation.

The catalytic SET domain of Whsc1 is dispensable for efficient pluripotency exit and mesendoderm differentiation. Whsc1 requires the catalytic SET domain on the carboxy (C) terminus to di-methylate the lysine 36 on histone 3 (Supplementary Fig. 4a)\(^29,30\). This function has been linked to the normal foetal heart and cartilage development in a murine model in which the SET domain of Whsc1 was excised\(^31\). Surprisingly, however, ESCs derived from these mice (hereafter referred to as ΔSET) behaved like wild-type ESCs, caused a delay in pluripotency exit, and EBs derived from these cells also showed retention of pluripotency gene expression and selective reduction of mesendoderm marker expression (Supplementary Fig. 4b,c). As the Whsc1 gene produces two major transcripts encoding a full-length protein of ~180 kD and a short isoform of ~100 kD that corresponds to the amino (N)-terminal
In contrast, the C-terminal construct containing the SET domain genes were found to be differentially expressed in Whsc1 the pluripotency genes, Pou5f1 ESCs caused by the knockout, we performed an RNA sequencing of either N-terminal construct (Nter-1 or -2) in phenotype in both assays (Fig. 4b,c). Remarkably, the expression layer formation assays; as expected, FL-WHSC1 rescued the cell's mesendoderm marker upregulation and pluripotency gene down-regulation of either N-terminal construct was also sufficient to restore, 

We found that the N-terminal domain of Whsc1 is sufficient for the biological properties of the protein in ESC differentiation, we performed rescue experiments. For this, we engineered Flag-tagged constructs of full-length human WHSC1 (FL-WHSC1), two N-terminal fragments of different lengths (Nter-1 and -2) and a C-terminal portion containing the SET domain (Fig. 4a). Whsc1<sup>1−/−</sup> cells expressing these constructs were subjected to the pluripotency exit and germ layer formation assays; as expected, FL-WHSC1 rescued the cell's phenotype in both assays (Fig. 4b,c). Remarkably, the expression of either N-terminal construct (Nter-1 or -2) in Whsc1<sup>1−/−</sup> cells was sufficient to rescue the cell's phenotype, showing rapid downregulation of Rex1, Pou5f1 and Nanog after induction of differentiation. In contrast, the C-terminal construct containing the SET domain failed to rescue (Fig. 4b). In line with these findings, the expression of either N-terminal construct was also sufficient to restore mesendoderm marker upregulation and pluripotency gene down-regulation during EB formation, while the C-terminal construct showed no such effect (Fig. 4c). Since Whsc1 has been described to methylate H3K36 and to alter EZH2 binding in multiple myeloma cells<sup>32,35</sup>, we compared Whsc1<sup>1−/+</sup> and Whsc1<sup>1−/−</sup> ESCs for the relevant histone marks. However, we found no significant differences in the global levels of H3K36me2, H3K36me3 and the EZH2-associated H3K27me3 mark (Supplementary Fig. 4e).

Our findings show that the N terminus of Whsc1 is sufficient to facilitate rapid exit from pluripotency and induce mesendoderm formation. They also suggest that the methyltransferase activity of Whsc1 acts in a cell context-dependent manner.

Whsc1 mediates enhancer activation of mesendoderm regulators. Our observations raise the possibility that Whsc1 directly controls the expression of transcription factors that specify mesendodermal and endodermal lineages. We therefore asked whether Whsc1 binds to and activates mesendodermal-specific enhancers. We focused on genomic loci of the mesendoderm instructive transcription factor genes Gata4, T, Gata6 and Foxa2. Putative enhancers of the neuroectodermal transcription factor genes Pax6 and Sox1 were included as controls. To identify candidate regulatory regions of these genes, we mapped the distribution of the enhancer mark H3K27ac in cells positive for either of the mesendodermal markers Eomes or Flk1, in mesendodermal progenitors induced by activin A, and in neural progenitor cells. In addition, as it has been shown that promoter–enhancer interactions occur mainly within the same topologically associating domains (TADs)<sup>16,14</sup>, we focused on regions within the same TAD. These analyses revealed the presence of several putative enhancers selectively marked by H3K27ac in either mesendodermal or neuroectodermal cells (Fig. 5a and Supplementary Fig. 5a).

To explore whether Whsc1 binds to these regions, we performed chromatin immunoprecipitation (ChIP) analyses using day 6 EBs. This showed that Whsc1 is significantly enriched at 8 out of 26 putative enhancers tested within the T, Gata6, Foxa2 and Gata4 loci and that this enrichment is reduced in Whsc1<sup>1−/−</sup> cells (Fig. 5b and Supplementary Fig. 5b). Notably, no significant Whsc1 binding was observed in 12 regions specifically marked by H3K27ac in the Pax6 and Sox1 loci in ectodermal cells (Supplementary Fig. 5b). To further explore whether these Whsc1-bound regulatory regions are required for target gene regulation, we performed UMI-4C experiments using promoter regions of T, Gata6, Foxa2 and Gata4 as viewpoints in ESCs and day 6 EBs, comparing Whsc1<sup>1−/−</sup> and Whsc1<sup>1−/−</sup> cells. We found that most Whsc1-bound putative mesendodermal enhancers appear to interact with target gene promoter regions in Whsc1<sup>1−/−</sup> day 6 EBs but not in ESCs, and that these contacts are lost/decreased in Whsc1<sup>1−/−</sup> day 6 EBs (Fig. 5c and Supplementary Fig. 5c). These observations are consistent with the notion that the relevant regions are involved in Whsc1-mediated target gene regulation.

To test whether Whsc1 binding is associated with enhancer activity, we monitored H3K27ac enrichment at Whsc1-bound regions in Whsc1<sup>1−/−</sup> and Whsc1<sup>1−/−</sup> day 6 EBs within the T, Gata6, Foxa2 and Gata4 loci. Regions not bound by Whsc1 served as controls. All ten Whsc1-bound regions showed a significant decrease of H3K27ac decoration in the Whsc1<sup>1−/−</sup> cells. In contrast, in the same cells, 14 out of 16 regions not bound by Whsc1, and all 12 sites within the Sox1 and Pax6 regions tested, showed no significant H3K27ac change (Fig. 5d and Supplementary Fig. 5d). Similar results were obtained for the H3K4me2 mark (Supplementary Fig. 5e). In addition, the ten Whsc1-bound regions showed no significant binding of Whsc1 in ESCs (Supplementary Fig. 5f) and their H3K27ac enrichment in ESCs was lower than in day 6 EBs (Supplementary Fig. 5f).

Taken together, these data show that Whsc1 binding is associated with active enhancers of target genes.

Whsc1 interacts with Brd4 to facilitate activation of mesendoderm enhancers. The bromodomain protein Brd4 has been reported to be required for enhancer activation and gene expression<sup>35</sup> and to interact with Whsc1 (ref. 36). We therefore tested whether Brd4 interacts with Whsc1 at the enhancer regions described above (Fig. 5b and Supplementary Fig. 5b). We found that in day 6 Whsc1<sup>1−/−</sup> EES Brd4 indeed binds to the enhancers of T, Gata6, Foxa2 and Gata4, as well as to those of Sox1 and Pax6. Importantly, Brd4 binding was significantly decreased in Whsc1<sup>1−/−</sup> EBS at the mesendoderm enhancers but not at the ectodermal enhancers (Fig. 6a). In addition, EBs treated during their formation with the Brd4 inhibitor JQ1 (ref. 37) showed reduced expression of the mesendodermal regulator genes Gata4, T, Gata6 and Foxa2, compared with dimethyl sulfoxide-treated control EBs (Fig. 6b).
Fig. 3 | Ablation of Whsc1 results in delayed downregulation of pluripotency genes and impairment of mesendoderm formation. a, Schematic of the strategy for generating ESCs with a complete Whsc1 knockout using the CRISPR–Cas9 approach. Top, organization of Whsc1, which can generate two transcript isoforms (long and short). The numbers above the schematic represent genomic localisation of Whsc1 gene. The targets of the gRNAs (exon 1 and exon 15) are indicated by arrows; Middle, genomic sequences of Whsc1 transcript isoforms (long and short). The numbers above the schematic represent genomic localisation of Whsc1, which can generate two ESC clones. b, Western blot of Whsc1 and the pluripotency factors Sox2, Oct4 and Nanog in Whsc1+/+ and two Whsc1−/− ESC clones. Tubulin was used as a loading control. n = 2 independent experiments were performed with similar results. Scanned images of unprocessed blots are shown in Supplementary Fig. 8. c, Expression kinetics of pluripotency genes in Whsc1+/+ and Whsc1−/− cells were monitored by RT-qPCR during exit of pluripotency by transferring into N2B27 medium containing activin A and FBS. d, Expression kinetics of mesodermal, ectodermal and pluripotency genes were evaluated by RT-qPCR during EB formation using Whsc1+/+ or Whsc1−/− ESCs (n = 3). In c and d, data represent means ± s.d. from n = 3 independent experiments, and P-values were calculated by two-tailed unpaired t-test. e, Volcano plot showing differentially expressed genes in Whsc1−/− ESCs compared with Whsc1+/+ ESCs, as determined by RNA-Seq of 8,283 genes. All genes with RPKM > 1 are plotted. Those upregulated in Whsc1−/− cells are indicated as red spots (log2[fold change] > 1 and P < 0.05), while those downregulated in Whsc1−/− cells are indicated as light blue spots (log2[fold change] < −1 and P < 0.05). The pluripotency genes Pou5f1, Sox2, Nanog, Esrrb and Rex1 are indicated as black spots.
We also observed an increase in the neuroectoderm markers Sox1, Pax6 and Nestin in JQ1-treated EBs, in line with the reported induction of neural differentiation after Brd4 depletion in ESCs \(^3\) (Fig. 6c). Moreover, both Whsc1 isoforms were found to co-immunoprecipitate with Brd4 in day 6 EB extracts (Fig. 6d), consistent with the notion that the N-terminus of Whsc1 is sufficient for this interaction.

Together, our results show that Whsc1 interacts with Brd4, and that this interaction favours the activation of mesendoderm enhancers but not of ectoderm enhancers.

**Gata6 recruits Whsc1 to a subset of mesendodermal enhancers.** Next, we asked whether mesendoderm-specific regulators are also able to interact with Whsc1 and recruit the factor to their target sites. As a model, we chose to study Gata6; first, we performed co-immunoprecipitation experiments. Indeed, both long and short Whsc1 isoforms co-immunoprecipitate with Gata6 in day 6 EBs, suggesting that the SET domain, which is not contained in the 100-kDa isoform, is dispensable for this interaction (Fig. 6e). To probe the second question, we generated Gata6 knockout ESCs by CRISPR–Cas9-mediated genome editing. These cells retain an
Fig. 5 | Whsc1 controls the enhancer activity of mesendoderm transcription factors. a, Top, HiC contact maps at 10-kb resolution from ESCs around T (left) and Gata6 (right) loci (GSE96611). TADs are marked by solid black lines, and sub-TADs are indicated by dashed black lines. The regions analysed using H3K27ac ChIP-Seq data are indicated as black boxes. Bottom, H3K27ac ChIP sequencing profiles on the loci of T (left) and Gata6 (right) in neural progenitor cells (NPCs) (GSE35496), Eomes+ mesendodermal progenitors (Eomes+)(GSE103262), activin A-induced mesendodermal precursors (MP) (GSE38596) and Flk1+ mesendodermal progenitors (Flk1+)(GSE47082). The black arrows below each panel correspond to putative regulatory regions up- or downstream of the respective gene. Mbp, megabase pairs. b, ChIP quantitative PCR (ChIP-qPCR) quantification of Whsc1 occupancy on the same enhancers as shown in a, in day 6 EBs from Whsc1+/+ and Whsc1−/− cells, with numbers indicating the distance to the transcription start site (TSS). c, Replicated UMI-4C profiles for baits located on the T (chromosome 17: 8,380,610–8,519,155) (left) and Gata6 (chromosome 18: 10,937,533–11,119,162) (right) promoters assayed in Whsc1+/+ and Whsc1−/− ESCs and day 6 EBs. Top, average contact profiles generated from the average of n = 2 independent biological replicates. Bottom, average contact fold change of day 6 EBs versus day 0 ESCs from n = 2 independent biological replicates. d, H3K27ac enrichments on putative enhancers of T (left) and Gata6 (right) were quantified by ChIP-qPCR in day 6 Whsc1+/+ and Whsc1−/− EBs. In b and d, data represent means ± s.d. from n = 3 independent experiments, and P values were calculated by two-tailed unpaired t-test.
ESC-like phenotype when grown in medium containing LIF and 2i. The Gata6−/− ESCs were then used to generate EBs and to perform Whscl ChIP experiments in day 4 EBs. We tested all ten previously identified Whscl-bound enhancer regions of T, Gata4, Gata6 and Foxa2 and found significantly decreased binding of Whscl at six of these regions (Fig. 6f).

These findings suggest that the mesendodermal lineage specifier Gata6 is capable of recruiting Whscl to a subset of mesendoderm-specific enhancers.

Depletion of mesendoderm transcription factors in ESCs causes a delay in pluripotency exit. The finding that Whscl is associated with the activation of Gata4, T, Gata6 and Foxa2 during ESC differentiation raised the possibility that the encoded mesendoderm instructive transcription factors are downstream mediators of Whscl in pluripotency exit and germ layer specification. To test this hypothesis, we transfected Rex1GFPd2 ESCs with small interfering RNAs (siRNAs) against Gata6, T, Foxa2 and Gata6, respectively, resulting in significant downregulation of the genes (Supplementary Fig. 6a). The cells were then induced to differentiate by transfer into N2B27 medium containing activin A and FBS. While control siRNA cells showed a complete loss of GFP expression at 48h postinduction, a significant proportion of cells transfected with siRNAs against Gata6, T, Gata4 and Foxa2 remained GFP positive (Supplementary Fig. 6b). To further study the role of these factors in pluripotency exit, 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 for Whscl−/− cells, Gata6−/− and Foxa2−/− cells showed a delayed downregulation of the pluripotency genes Oct4, Rex1 and Nanog (Fig. 7a). 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 72h (Fig. 7b). Moreover, when these cells were induced to generate EBs, they showed delayed downregulation of pluripotency markers similar to that found with Whscl−/− EBs (Fig. 7c).

In conclusion, our data show that the depletion or knockout in ESCs of the mesendoderm regulators Gata6, T, Gata4 and Foxa2 downstream of Whscl delays pluripotency exit.

Discussion

Here, we describe that the CRF Whscl has a role in both pluripotency exit and germ layer specification of ESCs. Depletion of Whscl entrapsthe cells in a pluripotent state and inhibits their specification towards mesoderm and endoderm lineages, without impairing neuroectoderm formation. Mechanistically, Whscl binds to and activates enhancers of mesendoderm lineage-instructive transcription factors during lineage specification, independent of its methyltransferase activity mediated by the SET domain. Moreover, depletion of downstream mesendodermal regulators also delayed pluripotency exit, explaining—at least in part—the effects observed after Whscl ablation.
Several factors previously described to be required for pluripotency exit have been shown to dismantle directly the pluripotency-associated regulatory network. Hence, folliculin, together with its interaction partners Fnip1 and Fnip2, sequesters the pluripotency-associated transcription factor Tfe3 in the cytoplasm; the transcription factors Tcf3, Foxd3 and Zfp706 repress pluripotency gene expression; and the RNA-binding proteins Mettl3 and Pum1 destabilize pluripotency gene transcripts. In contrast, we have found that Whsc1 does not act primarily on the pluripotency network, but controls the upregulation of mesendodermal regulators, including Gata4, T, Gata6 and Foxa2, which on their own also control pluripotency exit. Our observations suggest that the regulatory networks controlling pluripotency and germ layer specification are intimately linked. Indeed, several studies have shown that, during murine gastrulation in the primitive streak, the pluripotency factors Nanog, Oct4 and Sox2 are co-expressed with the mesendoderm lineage factors T, Gata6 and Foxa2. Moreover, it has been shown that pluripotency transcription factors are crucial for the upregulation of germ layer genes on differentiation, as differentiation fails when pluripotency transcription factors are acutely ablated. In addition, one study also suggested that, during in vitro differentiation of ESCs, GATA6 represses pluripotency factor expression.

Our findings support the notion that Whsc1 is recruited to chromatin by lineage regulators. Thus, co-immunoprecipitation experiments showed that Whsc1 interacts with Gata6 and that it loses its ability to bind to most of the mesendodermal enhancers in Gata6-deficient EBs. It is possible that other sites bound by Whsc1...
additional regulators are operative, such as has been described for Nkx2.5 in cardiac cells\(^1\). Our study also showed an interaction between Whsc1 and the transcriptional co-activator Brd4 in EBs, supporting earlier studies\(^2\). Moreover, EBs lacking Whsc1 showed a reduced binding of Brd4 specifically at mesendodermal enhancers, suggesting that Whsc1 facilitates the recruitment of Brd4 to these enhancers, leading to their activation. The finding that the BRD4 inhibition by the BET (bromodomain and extra-terminal motif) inhibitor JQ1 impairs the upregulation of mesendoderm regulators while dramatically enhancing the expression of ectodermal regulators supports previous observations in ESCs\(^3\). Together, our data suggest that Whsc1 enables Brd4 to act as a co-activator of mesendodermal enhancers.

In certain forms of cancer, including multiple myeloma and acute lymphoblastic leukaemia, WHSC1 has been found to be either over-expressed or hyper-activated, resulting in increased methylation of H3K36 on promoters of oncogenes that drive the disease\(^4\). This is mediated by the methyltransferase activity of the SET domain of WHSC1 described here is also involved in gene regulation during embry development, cardiogenesis and cancer.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41556-019-0342-1.

Received: 26 July 2017; Accepted: 9 May 2019; Published online: 24 June 2019

**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. Cleurs, 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, I. E. 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. Wray, I. 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).
9. Respuela, P. et al. Foxd3 promotes exit from naive pluripotency through enhancer decompensation and inhibits germline specification. *Cell Stem Cell* **18**, 118–133 (2016).
10. Martello, G. et al. Esr2b is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11**, 491–504 (2012).
11. Lee, 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).
43. Loh, K. M. & Lim, B. Epigenetics: actors in the cell reprogramming drama. Nature 488, 599–600 (2012).
44. 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).
45. Wamaitha, S. E. et al. Gata6 potently initiates reprogramming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. Genes Dev. 29, 1239–1255 (2015).
46. Jaffe, J. D. et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat. Genet. 45, 1386–1391 (2013).
47. Shen, C. et al. NSD3-short is an adaptor protein that couples BRD4 to the CHD8 chromatin remodeler. Mol. Cell 60, 847–859 (2015).
48. Bennett, R. L., Swaroop, A., Troche, C. & Licht, J. D. The role of nuclear receptor-binding SET domain family histone lysine methyltransferases in cancer. Cold Spring Harb. Perspect. Med. 7, a026708 (2017).
49. 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).
50. 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).

Acknowledgements
We thank A. Smith for providing the Rex1GFPd2 cells, K. Nimura for the Whsc1 ΔSET cells, E. J. Robertson for the Eomes-GFP cells, A. Surani for the Xcn1 EpiSCs, J. E. Bradner for the JQ1 compound, L. Batlle and C. Berenguer for technical help, the CRG core facilities of flow cytometry, advanced light microscopy and tissue engineering, and L. Di Croce, B. Payer, M. Behringer, J. Licht, K. M. Loh, K. Kaji 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 FJCI-2014-21872), B.D.S. by an EMBO long-term fellowship (number ALTF 1143-2015), G.S. by a Marie Sklodowska-Curie fellowship (H2020-MSCA-IF-2016, miRStem), A.D. by a Seveo Ochoa fellowship and J.G. by a Boehringer Ingelheim Graduate Student Fellowship. R.J. was supported by NIH grants R01 NS088538-01 and 2R01MH109610-15. This work was supported by the EU-FP7 project BLUEPRINT, the Spanish Ministry of Economy, Industry and Competitiveness to the EMBL partnership, Centro de Excelencia Severo Ochoa 2013–2017 and the CERCA Program Generalitat de Catalunya. T.V.T. and B.D.S. were supported by a CRG award for junior collaborative projects.

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., G.S., A.D., J.L.S., C.S.M., L.D.A.A. and J.G. performed the experiments. R.J. provided the reagents. E.V., M.V.-C. and A.G. conducted the bioinformatics analysis.

Competing interests
R.J. is an advisor/co-founder of Fate Therapeutics, Fulcrum Therapeutics, Omega Therapeutics and Dewpoint Therapeutics.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41556-019-0342-1.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.V.T. or T.G.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019
Methods

Cell culture. Mouse ESCs were cultured on plastic dishes precoated with 0.1% gelatin (EMD Millipore) with either DMEM/F12 medium and Neurobasal Medium mixed at a 1:1 ratio, supplemented with MEM Non-Essential Amino Acids solution (1×), sodium pyruvate (1 mM), l-glutamine (2 mM), penicillin (100 U ml⁻¹), streptomycin (100 μg ml⁻¹), β-mercaptoethanol (50 μM), N2 and B2 supplements, the small-molecule inhibitors PD0325901 (1 μM; Selleck Chemicals) and CHIR99021 (3 μM; Selleck Chemicals), and ESGRO LIF (1,000 U ml⁻¹; EMD Millipore), or with KnockOut DMEM medium supplemented with 15% (v/v) Embryonic Stem Cell FBS Qualified, MEM Non-Essential Amino Acids solution (1x), sodium pyruvate (1 mM), l-glutamine (2 mM), penicillin (100 U ml⁻¹), streptomycin (100 μg ml⁻¹), and 50 μM β-mercaptoethanol. Mouse EpiSCs (X²EpSC) were cultured on plastic dishes precoated with fibronectin (EMD Millipore) in DMEM/F12 medium, supplemented with MEM Non-Essential Amino Acids solution (1x), sodium pyruvate (1 mM), l-glutamine (2 mM), penicillin (100 U ml⁻¹), streptomycin (100 μg ml⁻¹), and 50 μM β-mercaptoethanol. Mouse EpiSCs (X²EpSC) were generated using a lentiviral system that facilitates genomic deletion by delivering simultaneously two gRNAs plus Cas9 into ESCs using Nucleofector according to the manufacturer's instructions. siRNA (GE Healthcare Dharmacon) (Supplementary Table 3) were electroporated into ESCs using Nucleofector according to the manufacturer's instructions. The pLKO-TRC-based shRNA constructs against Whsc1, Cbx4, Lmnb1 and shScr were purchased from Sigma–Aldrich (Supplementary Table 3). Lentiviral particles were produced by co-transfecting the constructs with pCMVΔ8.9 and pVSVG vectors in 293T cells. After 48 h, the supernatant was collected and filtered through a 0.22-μm filter, concentrated by ultracentrifugation at 80 °C for 90 min. Infected cells were then transferred to gelatin precoated dishes filled with complete medium including inhibitors and LIF. Three days later, infected cells were selected using puromycin (2 μg ml⁻¹) for an additional 3 d. Next, 20 pmol of ON-TARGETplus SMARTpool siRNAs against Gatdc1a, Gata2, Gatdc1b, and control sRNA (GE Healthcare Dharmacon) (Supplementary Table 3) were electroporated into ESCs using Nucleofector according to the manufacturer's instructions.

Genome editing with CRISPR-Cas9. The two guide RNAs (gRNAs) targeting the Whsc1 locus were designed using the CRISPR design tool from MIT (http://crispr.mit.edu) (Supplementary Table 3). After annealing, the two gRNAs were cloned into the DECKO (Double Excision CRISPR Knockout) version of SpCas9 (BB; 2A-GFP) (PX438) (original backbone from Addgene number 48130) [55]. The DECKO system facilitates genomic deletion by delivering simultaneously two gRNAs plus Cas9 into E14 cells [56]. Then, 24 h after transfection, single cells were sorted by fluorescence-activated cell sorting. Mutant clones were screened through PCR and sequencing to identify homology-targeted clones.

Alkaline phosphatase assay. Alkaline phosphatase assays were performed using the Alkaline Phosphatase Staining Kit II (Stemgent) following the instructions provided by the manufacturer.

Assay for pluripotency exit. ESCs were dissociated with Accutase for 2 min at room temperature and washed twice with PBS. Then, 2 × 10⁵ cells were seeded in 6-well plates precoated with 0.1% (w/v) EmbryoMax gelatin in culture medium containing N2B27 supplements plus activin A (25 ng ml⁻¹; PeproTech) and 10% FBS without the two inhibitors or LIF. GFP expression levels were monitored by flow cytometry on an LSRFortessa analyser (BD Biosciences), and the analysis was performed using FlowJo software version 10.6.5 (TreeStar). In some experiments, Rxi1GFP2 cells that retained GFP expression were selected by adding 10 μg ml⁻¹ blastocidin (Sigma–Aldrich) to the medium.

EB formation, and cardiac progenitor, definitive endoderm and neural progenitor differentiation. ESCs or EpiSCs were induced to form EBs with 'EB differentiation medium' (KnockOut DMEM medium supplemented with 10% FBS, MEM Non-Essential Amino Acids solution (1x), sodium pyruvate (1 mM), l-glutamine (2 mM), penicillin (100 U ml⁻¹), streptomycin (100 μg ml⁻¹), and 50 μM β-mercaptoethanol and plated as hanging drops (400 cells in 20 μl) on the lid of a 100-mm Petri dish. The cultures were harvested 3 d later and transferred into bacterial dishes for further culture in suspension. The cardiac progenitor differentiation assay was performed as described [57] 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 gelatin precoated 96-well plate supplemented with 100 μl EB differentiation medium and cultured for another 3 d before counting the beating EBs under low-power magnification. The definitive endoderm and neural differentiation protocols were performed as described previously [58].

Animal experiments and teratoma formation. Teratoma formation assays were performed as described previously [59]. Briefly, knockdown and control ESCs grown under 2i plus LIF conditions (0.5 μM iodoacetamid) were injected subcutaneously into severe combined immunodeficient 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 haematoxylin and eosin for histological analysis. All experimental procedures were approved by the local ethical committee (Comité Ético de Experimentación Animal del Parc de Recerca Biomédica de Barcelona), and met the guidelines of the local (Catalan law 5/1995 and Decrees 214/97 and 32/2007) and European regulations (EU directives 86/609 and 2001–486) as well as the Standards for Use of Laboratory Animals A3588-01 (National Institutes of Health).

Imaging acquisition and analysis. Fluorescence images were acquired using Leica DMi6000 B fluorescence microscope. Images of teratoma haematoxylin and eosin staining and EB morphology images were acquired using Leica DMi6000 B light microscopy. The Leica application software LAS AF was used to visualize the images. Image J (version 1.15) was used for further analysis.

ChIP assay. ChIP assays in ESCs and EBs were performed using a True MicroChIP Kit (Diagenode) according to the manufacturer's instructions. EBs were cross-linked with 4% formaldehyde for 20 min for H3K27Ac and H3K4me2, or double-cross-linked with DSG (disuccinimidyl glutarate) 2 μM for 30 min, then with 4% formaldehyde for 20 min, for Brd4 and Whsc1. Genomic DNA was purified using MicroChIP DiaPure columns. The antibodies and primers used are listed in Supplementary Table 3.

Western blot and co-immunoprecipitation analysis. Proteins were extracted using NET-2 buffer (50 mM Tris (pH 7.4), 200 mM NaCl, 0.1% triton and proteases inhibitors) and loaded on sodium dodecyl sulfate polyacrylamide gel electrophoresis gel. For the endogenous co-immunoprecipitation experiments, lysates were prepared using NET-2 buffer, and centrifuged to eliminate insoluble material. The extract was precleared by incubation with magnetic beads for 2 h on a rotating wheel at 4 °C. To crosslink the antibodies to Dynabeads A or G (Life Technologies) magnetic beads, 50 μl of beads were incubated with 3 μg of antibody or immunoglobulin G for 2 h at room temperature on a wheel. The mix was then washed once with PBS and twice with triethanolamine (pH 8.2; 0.2 M) and then incubated for 20 min at room temperature with dimethyl pimelidate (20 μM). To stop the reaction, two 5 min washes with Tris-Cl (pH 8; 50 mM) were performed, followed by three washes with PBS. Then, beads were incubated with citrullinated peptides (1 μM; pH 3) for 2 min. Finally, the 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 was washed six times with lysis buffer. To elute, the last was washed and the beads resuspended in 1x Laemmli buffer without β-mercaptoethanol and heated at 100 °C. Magnetic beads were then separated and supernatant was taken, complemented with 5% β-mercaptoethanol and boiled before loading on a sodium dodecyl sulfate polyacrylamide gel to analyse the bound proteins by western blotting.

Gene expression analysis. Total RNA was isolated using an RNaseasy Mini Kit (Qiagen) according to the manufacturer's instructions. Double-stranded complementary DNA (cDNA) was synthesized using the High Capacity RNA-to-cDNA Kit (Applied Biosystems). Real-time PCR was performed on a Viia 7 system using Power SYBR Green PCR Master Mix (Applied Biosystems). The relative gene expression levels were calculated using the 2⁻ΔΔCt method. Primers can be found in Supplementary Table 3. RNA-Seq libraries were prepared using the Illumina TruSeq Stranded mRNA Library Prep Kit followed by single-end sequencing (50 bp) on an Illumina HiSeq 2500. Single-end reads were mapped using STAR2.5 (ref. 60). Gene expression was quantified using STAR (–quantMode GeneCounts) and genes with RPKM values of <0.1 in at least two samples were removed. Genes with significant differential expression were identified with a t-test (P < 0.05) and an absolute fold change of ≥2.

Chromosome conformation capture followed by high-throughput sequencing data generation and analysis. Two independent replicates of Whsc1 and Whsc1 embryonic stem cells and day 6 EBs were prepared following the protocol described [61]. Briefly, two million cells were cross-linked using 1% formaldehyde and
quenched using 0.125 M glycine. The cell pellet was lysed in prepared ice-cold HiC lysis buffer (10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% IGEPAL CA-630 and 1× Roche complete protease inhibitors). The chromatin was then digested using MboI (NEB R147M) at 37°C in four rounds as follows: 100 U for 2 h, 100 U for 2 h, 100 U overnight and 100 U for 2 h. An aliquot was taken from the tube and analysed by gel electrophoresis to inspect digestion efficiency. If digestion was successful, the restriction enzyme was inactivated by incubating at 65°C for 20 min. Next, the chromatin was ligated with 10,000 U T4 ligase (NEB M0202M) with 10× ligase buffer to a final volume of 1.5 mL and incubated overnight at 24°C. Ligation efficiency was inspected by gel electrophoresis on an aliquot of the sample. The chromatin was then reverse cross-linked for 1 h at 55°C with 50 μl proteinase K (Sigma–Aldrich; 20 mg/mL) and 10 μl RNase A (Sigma–Aldrich; R4642) for 4 h at 65°C. The 3C DNA was then extracted with ethanol precipitation and resuspended in Tris buffer (10 mM Tris-HCl, pH 8). 3C aliquots were stored at −20°C for later use. 4C primers were designed flowing a nested approach involving a set of two primers for each locus, thereby improving specificity. The primer for the second PCR reaction should ideally be located 5–15 bp from the interrogated restriction site. The primer used in the first PCR reaction is designed upstream of the second primer with no overlap. For multiplexing, we designed primers independently, aiming for a melting temperature (Tm) of 60–65°C. A list of the primers used in this study is provided in Supplementary Table 3. Some 3 μg of DNA in 300 μL Tris buffer (10 mM Tris-HCl, pH 7.8) of 3C sample was sonicated at 4°C for seven cycles of 30 s on, 60 s off on a Bioruptor (Diagenode). The sonicated material was inspected by gel electrophoresis of 5 μl aliquots. The sonicated DNA was subjected to an end-repair reaction—10 μl 10× end-repair buffer and 5 μl end-repair mix (NEB E6050L) were added to the sonicated DNA and incubated at 20°C for 30 min. The DNA was cleaned with 1.1× AMPure XP beads (Beckman Coulter) and eluted in 42 μl Tris buffer. A-tailing was performed by adding 5 μl NEBNext dA-Tailing Reaction Buffer and 3 μl Klenow Fragment (NEB M0212M). Following A-tailing, the 5′ ends of the DNA were dephosphorylated by adding 2 μl calf intestinal alkaline phosphatase (NEB M0290S), and incubated for 30 min at 37°C and 60 min at 50°C. The DNA was cleaned with 1.1× AMPure XP beads and eluted in 35 μl elution buffer (10 mM Tris-Cl). The DNA template was ligated with 3 μl NEBNext Adaptor, 2 μl Quick T4 DNA Ligate and 10 μl 5× Quick Ligation Reaction Buffer for 15 min at 20°C. Then, 3 μl of USER Enzyme was added and incubated at 37°C for 15 min. To release the non-ligated strand of the adaptor, the DNA was denatured at 95°C for 2 min and cleaned with 1× AMPure XP beads. Two nested PCR reactions were used for 4C library construction. The PCR reactions were performed in a 50 μl volume with the following reagents: 10 μl Herculase II 5× buffer, 0.25 mM dNTPs, 3.75 μM 10× molecular length universal primer (Nextera; NEB), 1μl Herculase II polymerase and 10 μl DNA template. The PCR programme was as follows: 2 min at 95°C, 15 cycles of 30 s at 95°C, 30 s at 60°C and 60 s at 72°C, and final extension of 5 min in 72°C. Ten percent of the first PCR product was used as a template in the second PCR reaction. Conditions for the second PCR reaction were identical to the first; the only difference was the use of the nested-bait primer. PCR products for each library were purified with a 1:1× AMPure XP bead. The libraries were then pooled and diluted to a concentration of 4 nM, multiplexed with other libraries and sequenced on an Illumina HiSeq 2500.

For data analysis, we first preprocessed the reads, splitting them according to the restriction enzyme motif, and obtaining read fragments. Then, we aligned them against the mm10 genome using BW A (0.7.17; r1188) with all default arguments. The RNA-Seq and UMI-4C data generated have been deposited in the Gene Expression Omnibus under accession number GSE126618. The accession numbers of the published datasets used are as follows: HiC in ESCs: GSE96611 (ref. 40); H3K27ac Chip-Seq in Flk1+ mesendodermal progenitors: GSE47082 (ref. 42); H3K27ac Chip-Seq in Eomes+ mesendodermal progenitors: GSE103262 (ref. 43); H3K27ac Chip-Seq in activin A-induced mesendodermal precursors: GSE38596 (ref. 41); and KHC27ac ChIP-Seq in neural progenitor cells: GSE35496 (ref. 44). Source data for all figures have been provided as Supplementary Table 4. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

References

51. Gillich, A. et al. Epiblast stem cell-based system reveals reprogramming of germline factors. Cell Stem Cell 10, 425–439 (2012).
52. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).
53. Aparicio-Prat, E. et al. DECKO: single-oligo, dual-CRISPR deletion of genomic elements including long non-coding RNAs. BMC Genomics 16, 846 (2015).
54. Di Stefano, B. et al. C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature 506, 235–239 (2014).
55. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675 (2012).
56. Dobin, A. et al. STAR: ultrafast universal RNA-Seq aligner. Bioinformatics 29, 15–21 (2013).
57. Schwartzman, O. et al. UMI-4C for quantitative and targeted chromosomal contact profiling. Nat. Methods 13, 685–691 (2016).
58. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3999 (2013).
59. Bates, D. Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. Preprint at https://arxiv.org/abs/1406.5823 (2014).
60. R Development Core Team R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2018).
61. Stadhouders, R. et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. 50, 238–249 (2018).
62. Org, T. et al. Sci binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. EMBO J. 34, 759–777 (2015).
63. Alexanian, M. et al. A transcribed enhancer dictates mesendoderm specification in pluripotency. Nat. Commun. 8, 1806 (2017).
64. Yu, P. et al. Spatiotemporal clustering of the epigenome reveals rules of dynamic gene regulation. Genome Res. 23, 352–364 (2013).
65. Lodato, M. A. et al. SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS Genet. 9, e1003288 (2013).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] The statistical test(s) used AND whether they are one- or two-sided
- [ ] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [ ] A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [ ] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
- [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [ ] Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

FACS data were collected using FACS Diva V6.1.2 (BD Bioscience). qPCR data were collected using ViiA7 software. Leica application software LAS AF were used to visualize and collect images.

Data analysis

Statistical analysis was performed using PRISM 6 software. The p values were calculated using either two-tailed unpaired Student’s t test or two-tailed Mann-Whitney test (Supplementary Table 4).

FACS data were analyzed using Diva v6.1.2 (BD Biosciences) and FlowJo software v10.0.6 (TreeStar).

Image J (Version1.15) was used for imaging analysis.

For RNA-seq

Single-end reads were mapped using STAR2.5. Gene expression was quantified using STAR (–quantMode GeneCounts) and the Gencode mouse Release M17. Sample scaling and statistical analysis were performed using R3.3.2. RPKM values were used as an absolute measure of gene expression, and genes with RPKM values less than 1.0 in at least two samples were removed. Genes with significant differential expression were identified with a t-test (P Value<0.05) and absolute Fold Change >2.

For UMI-4C

We first pre-processed the reads splitting them according to the restriction enzyme (RE) motif and obtaining read fragments. Then we aligned them against the mm10 genome using bwa (0.7.17-r1188) [1] with all default arguments. We removed all duplicate and supplementary alignments (sam flag >= 2048) as well as those with a mapping quality below 20. For each bait, we summarized the information per restriction fragment, computing the tally of read fragments and obtaining a raw contact profile. We defined the region of interest as all restriction fragments within 2 Kbp and 250 Kbp distance to the bait. In order to smooth the profiles, we increased the size of each restriction fragment aggregating the adjacent restriction fragments sequentially until a threshold in contacts was reached (5% of the total number of raw contacts in the region of interest). We used a linear mixed model [2,3] to integrate replicates and compare conditions. In detail: We modeled the log smoothed contacts per restriction fragment assuming a linear decay with the log distance to the bait, different at each side of it. We included both random intercepts and slopes on the restriction fragment level (the slope defined at the condition to compare). From this model we obtained point and 95% CI estimates for 1) The average profile per condition 2) The fold change between them (corresponding to the random slopes of the restriction fragments).

[1] Li, Heng. “Aligning sequence reads, clone sequences and assembly contigs with...

[2]...

[3]...
RNA-seq and UMI-4C data generated has been deposited in the Gene Expression Omnibus (GEO) under accession number GSE126618. Accession number of published datasets used: HiC in ESCs GSE96611; H3K27ac ChIP-seq in Flk1+ mesendodermal progenitors: GSE47082; H3K27ac ChIP-seq in Eomes+ mesendodermal progenitors: GSE103262; H3K27ac ChIP-seq in Activin A-induced mesendodermal precursors: GSE38596; H3K27ac ChIP-seq in neural progenitor cells: GSE35496. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

- Sample size: All experiments were obtained from three or two independent experiments with similar outcomes. Teratoma experiments were performed with n=10 SCID mice. No sample size calculation was performed to predetermine sample size. All sample sizes are listed in the corresponding figure legend.
- Data exclusions: There was no samples or animals excluded from the analysis.
- Replication: All data presented were obtained from three or two independent experiments with similar outcomes. RNA-seq and 4C experiments were performed using two biological replicates. To assess the reproducibility of RNA-seq and UMI-4C experiments, we calculated pair-wise Spearman correlation coefficients between the individual samples (all >0.95).
- Randomization: Twenty SCID mice were randomized into two groups (10/group) and were injected with shScr or shWhsc1 cells to generate teratomas.
- Blinding: During data collection and analysis, the investigators were not blinded, because the results were directly collected by instrumentation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a | n/a |
| [x] Antibodies | [x] ChIP-seq |
| [ ] Eukaryotic cell lines | [x] Flow cytometry |
| [x] Palaeontology | [ ] MRI-based neuroimaging |
| [x] Animals and other organisms | |
| [x] Human research participants | |
| [ ] Clinical data | |

Antibodies

Antibodies used: Whsc1 antibody: ChIP 2ug, IP 5ug, WB 1:500, Abcam Cat# 75359, Clone 29D1
**Eukaryotic cell lines**

**Policy information about cell lines**

**Cell line source(s)**

293T and E14TG2a cells were from ATCC; Rex1GFPd2 and EomesGFP reporter cell lines was a kind gifts from A. Smith and E. J. Robertson; XGFP EpiSCs were kindly provided by A. Surani; Whsc1ΔSET ESCs were kindly provided by K. Nimura; Whsc1-/- cells were generated in our laboratory using CRISPR-Cas9 technology and verified by sequencing. Gata6-/- and Foxa2-/- were generated using CRISPR-Cas9 technology by Johanna Goldmann in Rudolf Jaenisch laboratory.

**Authentication**

293T and E14TG2a cells were from ATCC and have not been authenticated by STR profiling. Rex1GFPd2 and EomesGFP reporter cell lines have been verified in experimental conditions (Fig. 1 and Supplementary Fig. 3). Whsc1ΔSET ESCs were verified by Western blot (Supplementary Fig. 4). Whsc1ΔSET ESCs were verified by PCR and Western blot (Fig. 3). Gata6-/- and Foxa2-/- were verified by sequencing by Johanna Goldmann in Rudolf Jaenisch laboratory.

**Mycoplasma contamination**

All cell lines were regularly tested negative from mycoplasma.

**Commonly misidentified lines**

No commonly misidentified cell lines were used in this study.

---

**Animals and other organisms**

**Policy information about studies involving animals**

**ARRIVE guidelines** recommended for reporting animal research

**Laboratory animals**

SCID mice were from Charles River. Six-week old male mice were used to generate teratomas.

**Wild animals**

This study did not involve wild animals.

**Field-collected samples**

This study did not involve field-collected samples.

**Ethics oversight**

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 AS389-01 (NIH).

Note that full information on the approval of the study protocol must also be provided in the manuscript.
## Flow Cytometry

### Plots

Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

| Methodology | Sample preparation | Instrument | Software | Cell population abundance | Gating strategy |
|-------------|--------------------|------------|----------|---------------------------|-----------------|
|             | Rex1GFPd2 cells or EomesGFP reporter cells were dissociated with accutase then washed twice with PBS 1X, and stained with DAPI before FACS analysis. | GFP expression level was monitored by flow cytometry on a LSR Fortessa analyser (BD Bioscience). | The analysis was performed using Diva v6.1.2 (BD Biosciences) and FlowJo software v10.0.6 (TreeStar). | Purity of cell analyzed was determined by independent immunofluorescence imaging. E14TG2a cells (GFP-) were routinely used as negative control to gate GFP positive population. | We used FSC and SSC gates to include at least 90% of events and to exclude very small events. Then we use SSC-A and -H to discriminate for single cells (doublets were excluded for further analysis). Third, DAPI were used to discriminate live cells from dead ones. GFP and RFP gating was performed using negative controls from the respective cell lines tested. Example of gating strategy was shown in Supplementary Fig. 7 |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.