© 2016. Published by The Company of Biologists Ltd | Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

Sall4 controls differentiation of pluripotent cells independently of the Nucleosome Remodelling and Deacetylation (NuRD) complex

Anzy Miller[1,2], Meryem Ralser[1], Susan L. Kloet[3], Remco Loos[1,4], Ryuichi Nishinakamura[5], Paul Bertone[1,4], Michiel Vermeulen[3] and Brian Hendrich[1,2,*]

ABSTRACT

Sall4 is an essential transcription factor for early mammalian development and is frequently overexpressed in cancer. Although it is reported to play an important role in embryonic stem cell (ESC) self-renewal, whether it is an essential pluripotency factor has been disputed. Here, we show that Sall4 is dispensable for mouse ESC pluripotency. Sall4 is an enhancer-binding protein that prevents precocious activation of the neural gene expression programme in ESCs but is not required for maintenance of the pluripotency gene regulatory network. Although a proportion of Sall4 protein physically associates with the Nucleosome Remodelling and Deacetylase (NuRD) complex, Sall4 neither recruits NuRD to chromatin nor influences transcription via NuRD; rather, free Sall4 protein regulates transcription independently of NuRD. We propose a model whereby enhancer binding by Sall4 and other pluripotency-associated transcription factors is responsible for maintaining the balance between transcriptional programmes in pluripotent cells.

KEY WORDS: Sall4, NuRD, ES cells, Enhancer, Transcription factor, Co-repressor

INTRODUCTION

Embryonic stem cells (ESCs) have the potential to form any somatic cell type in the adult organism; that is, they are pluripotent. In order to properly execute lineage decisions, pluripotent cells must precisely coordinate their gene expression programmes. To successfully initiate differentiation down one particular lineage, a cell must activate the gene regulatory network (GRN) appropriate to enter that lineage, and not those corresponding to any other lineage, while also extinguishing the pluripotency GRN. It is clear from a large number of studies that the coordinated action of multiple transcription factors and chromatin-modifying proteins is essential to maintain the delicate balance between self-renewal and differentiation of ESCs (Morey et al., 2015; Niwa, 2007; Signolet and Hendrich, 2015). Although it is relatively straightforward to show that a given protein plays some role in ESC differentiation, often the precise mechanisms of how the important transcription factors function remain ill-defined.

In this study we focus on Sall1 and Sall4, the only two members of the *spalt* gene family of C2H2-type zinc-finger transcription factors that are expressed in ESCs (reviewed by de Celis and Barrio, 2009). In humans, mutations in *SALL4* show haploinsufficiency, resulting in the autosomal dominant Okihiro/Duane-Radial Ray and IVIC syndromes (Al-Baradie et al., 2002; Kohlhase et al., 2002; Sweetman and Munsterberg, 2006), while mutations in *SALL1* lead to the autosomal dominant Townes-Brocks syndrome (Kohlhase et al., 1998). *SALL4* is also aberrantly expressed in many cancers and correlates with poor prognosis, leading it to be heralded as a new cancer biomarker and potential therapeutic target (Zhang et al., 2015). In mice, Sall4 has been shown to play an essential role in peri-implantation development (Elling et al., 2006; Sakaki-Yumoto et al., 2006; Warren et al., 2007), while Sall1 is dispensable for early embryogenesis but is essential for kidney development (Kanda et al., 2014; Nishinakamura et al., 2001).

The role played by Sall4 in ESCs has been the subject of some debate. Studies using *Sall4* null ESCs concluded that it was dispensable for self-renewal of ESCs, but that mutant cells were prone to differentiate in certain conditions, indicating that it might function to stabilise the pluripotent state (Sakaki-Yumoto et al., 2006; Tsubooka et al., 2009; Yuri et al., 2009). By contrast, studies in which Sall4 was knocked down in ESCs led to the conclusion that it plays an important role in the maintenance of ESC self-renewal (Rao et al., 2010; Zhang et al., 2006). Sall4 was found to bind regulatory regions of important pluripotency genes such as of *Pou5f1* (previously known as *Oct4*) and *Nanog* (Wu et al., 2006; Zhang et al., 2006) and a physical interaction with the Pou5f1 and Nanog proteins has been reported (Pardo et al., 2010; Rao et al., 2010; van den Berg et al., 2010; Wu et al., 2006). The consensus arising from these studies was that Sall4 is instrumental in the regulation of key pluripotency genes and is thus a key regulator of the pluripotency transcriptional network (van den Berg et al., 2010; Xiong, 2014; Yang et al., 2010). Whether it is essential for self-renewal remains a point of contention.

Sall1 and Sall4 have both been shown to interact biochemically with the Nucleosome Remodelling and Deacetylase (NuRD) complex. NuRD is a transcriptional regulatory complex that has nucleosome remodelling activity due to the Chd4 helicase and protein deacetylase activity due to Hdac1 and Hdac2. Additional NuRD components are the zinc-finger proteins Gatad2a/b, SANT domain proteins Mta1/2/3, histone chaperones Rbbp4/7, structural protein Mbd3 (which can be substituted for by the methyl-CpG-binding protein Mbd2) and the small Cdk2ap1 protein (Allen et al., 2013; Le Guezennec et al., 2006). The usual interpretation of the

[1]Wellcome Trust – Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge CB2 1QR, UK. [2]Department of Biochemistry, University of Cambridge, Cambridge CB2 1QR, UK. [3]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, 6525 GA Nijmegen, The Netherlands. [4]European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. [5]Department of Kidney Development, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto 860-0811, Japan.

*Author for correspondence (brian.hendrich@cscr.cam.ac.uk)

B.H., 0000-0002-0231-3073

Received 26 April 2016; Accepted 18 July 2016

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

Sall-NuRD interaction is that Sall proteins recruit NuRD to influence transcription of their target genes (Kiefer et al., 2002; Kloet et al., 2015; Lauberth and Rauchman, 2006; Lu et al., 2009; Yuri et al., 2009). The relationship between Sall proteins and NuRD might not be so straightforward, however, as they show opposing functions in ESCs. Whereas Sall1 and Sall4 are implicated in maintenance of the ESC state, NuRD functions to facilitate lineage commitment of ESCs (Kaji et al., 2006; Reynolds et al., 2012; Signolet and Hendrich, 2015).

In this study we set out to define the function of Sall4 in ESCs and to understand the relationship between NuRD and Sall4. We use defined culture conditions (2i/LIF) (Ying et al., 2008) to show that Sall1 and Sall4 prevent activation of neural genes in ESCs, but are dispensable for the maintenance of the pluripotency GRN. We further show that although NuRD is the major biochemical interactor of Sall4, only ∼10% of Sall4 protein associates with NuRD in ESCs. Despite this interaction, Sall4 neither recruits the NuRD complex to chromatin nor shows NuRD-dependent transcriptional regulation. The majority of Sall4 has no stable biochemical interactors, but colocalises with pluripotency-associated transcription factors at enhancer sequences. We propose a model to explain why accumulation of these transcription factors can stimulate the transcription of some genes but inhibit the expression of others.

RESULTS

Sall4 is dispensable for ESC self-renewal, but inhibits neural differentiation

To define the function of Sall4 in pluripotent cells, ESCs were made homozygous for a previously described *Sall4* conditional allele (Sakaki-Yumoto et al., 2006) by two different methods: gene targeting and derivation from homozygous *Sall4* floxed mice followed by Cre-mediated recombination. The *Sall4* null ESC lines lack exons two and three, which contain all of the zinc-finger domains found in Sall4 (Fig. 1A). Although a truncated *Sall4* transcript is produced from this allele, no protein is detectable (Fig. S1A,B). To rule out potential compensation by the related Sall1 protein (Yuri et al., 2009), which is the only other Sall protein expressed in wild-type (WT) ESCs (Fig. S1A), we also derived ESCs from $Sall1^{flox/flox}$; $Sall4^{flox/flox}$ mice. These cells were then used to obtain $Sall1^{-/-}$; $Sall4^{flox/-}$ (referred to as *Sall1* null) and $Sall1^{-/-}$; $Sall4^{-/-}$ (referred to as *Sall4/1* double-null) ESC lines (Fig. 1A,B) after Cre transfection and clonal isolation. Deletion of either *Sall1* or *Sall4* had no effect on the transcription level of the other gene (Fig. S1A). *Sall1* null, *Sall4* null, and *Sall4/1* double-null ESCs were viable and were able to be maintained as self-renewing cultures in 2i/LIF conditions (Fig. S1C). All ESC lines tested (WT, *Sall1* null, *Sall4* null and *Sall4/1* double-null cells) were able to give rise to tissues representing all three germ layers in teratoma assays, indicating that Sall4 and Sall1 are dispensable for ESC potency (Fig. 1C).

Although loss of both Sall1 and Sall4 was compatible with self-renewal in 2i/LIF conditions, there was considerably more spontaneous differentiation in double-mutant cultures than with either single mutant. The *Sall4/1* double-null differentiated cells present in 2i/LIF cultures sent out long processes that stained positively for the neuronal marker TuJ1 (also known as Tubb3), indicative of postmitotic neurons (Fig. 1D). When plated into serum/LIF conditions (in the absence of feeders), both *Sall4* null and *Sall4/1* double-null cells showed widespread differentiation (Fig. S1D). By contrast, *Sall1* null cells behaved similarly to WT in all conditions tested in this study.

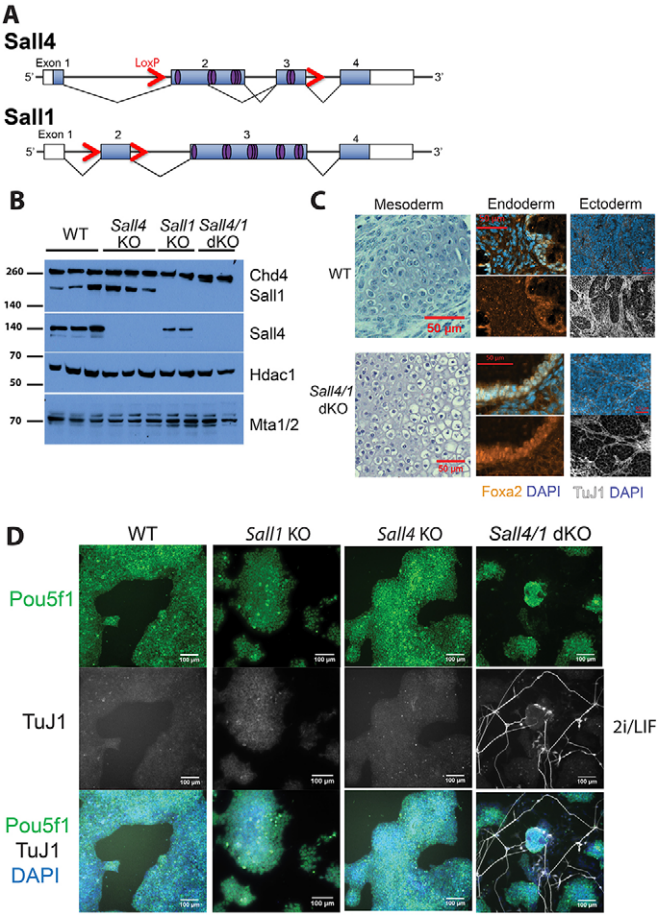


Fig. 1. Sall4 and Sall1 are dispensable for mouse ESC self-renewal. (A) Schematic of targeted *Sall4* and *Sall1* genomic loci. Boxes represent exons, and filled boxes indicate the coding sequence; red arrows represent LoxP sites; purple ovals represent zinc-finger domains. (B) Western blot of wild-type (WT), *Sall4* null (hereafter *Sall4* KO), *Sall1* null (hereafter *Sall1* KO) and *Sall4/1* double-null (hereafter *Sall4/1* dKO) ESC lines in 2i/LIF. The blot was probed with anti-Chd4, anti-Sall1, anti-Sall4, anti-Hdac1 and anti-Mta1/2 antibodies. Molecular weights are shown at left in kDa. Note that the *Sall1* KO line is heterozygous for *Sall4* ($Sall1^{-/-}$; $Sall4^{flox/-}$). (C) Representative images from teratoma assays of mesoderm, endoderm and ectoderm tissues derived from WT (top) and *Sall4/1* dKO. Cartilage tissue is shown for mesoderm by H&E staining, immunofluorescence staining for Foxa2 (with DAPI) for endoderm, and for TuJ1 (with DAPI) for ectoderm. (D) Immunofluorescence of WT, *Sall1* KO, *Sall4* KO and *Sall4/1* dKO ESCs grown in 2i/LIF, stained for Pou5f1 (green), TuJ1 (white) and with DAPI (blue). Out of all DAPI-stained *Sall4/1* dKO ESCs per field, 2.09±1.19 (mean±s.d.) also stained positively for TuJ1. Six images were used to generate counts. Scale bars: 50 µm in C; 100 µm in D.

These observations suggested that Sall4 and Sall1 are involved in suppressing neural differentiation in ESCs. To test this hypothesis, single- and double-mutant cultures were subjected to a standard neuroectodermal differentiation protocol (Ying et al., 2003). Whereas WT cultures did not produce TuJ1-expressing neurons during the first 5 days of this protocol, TuJ1-expressing cells displaying neuronal morphology could clearly be seen by day 5 in *Sall4* null cultures and by day 2 in the *Sall4/1* double-null cultures (Fig. 2A). After only 2 days of the protocol, the majority of *Sall4/1* double-null cells had activated expression of the neural progenitor marker Sox1, and many had extinguished Pou5f1 expression, whereas Pou5f1 was still ubiquitously expressed in WT cells at this point and only a few WT cells had activated Sox1 (Fig. 2B,C). Thus, absence of Sall proteins in ESCs results in an accelerated pace of

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

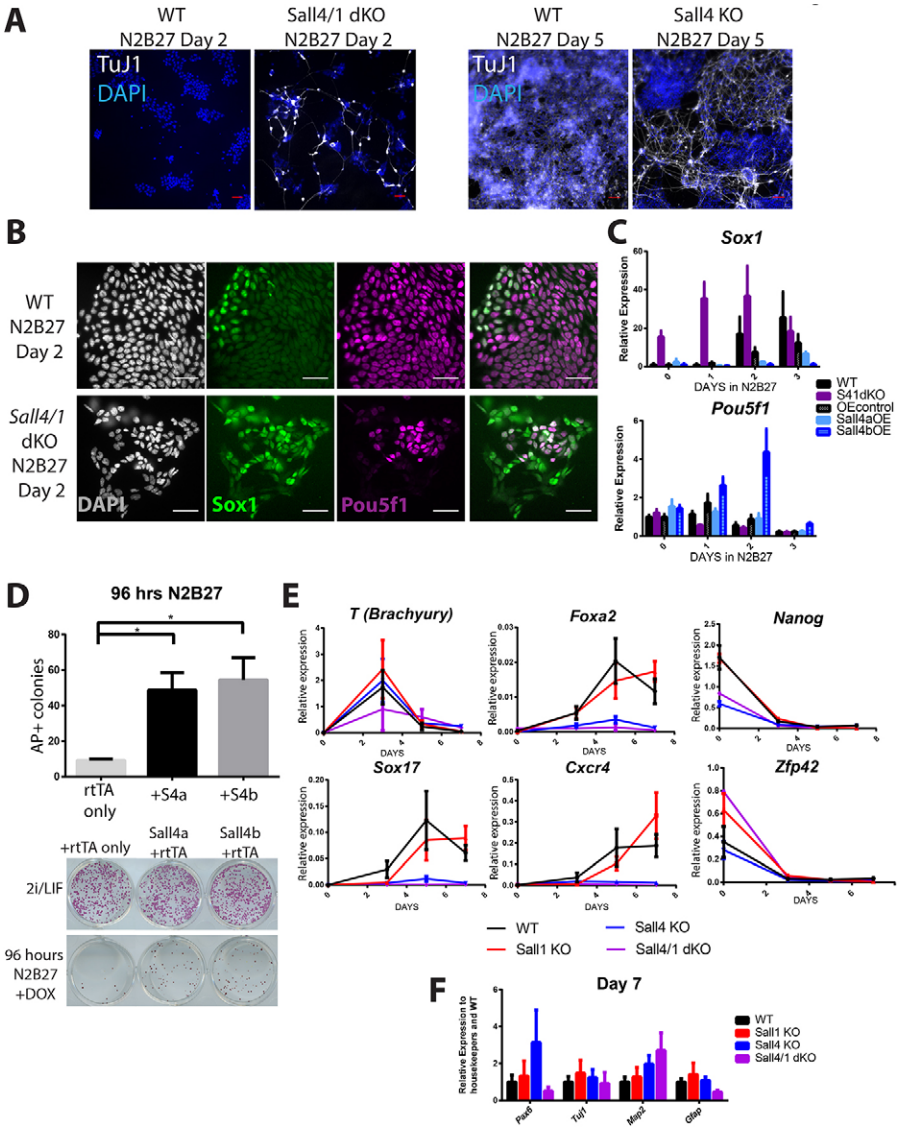


Fig. 2. Sall4 and Sall1 block neural differentiation. (A) Representative immunofluorescence images of WT and *Sall4/1* dKO cells after 2 days in N2B27 (left), or WT and *Sall4* KO after 5 days in N2B27 (right) stained for TuJ1 (white) and with DAPI (blue). (B) Representative immunofluorescence images of WT and *Sall4/1* dKO cells after 2 days in N2B27, stained with DAPI (white) or for Sox1 (green) and Pou5f1 (magenta). The right-hand image is a composite of Pou5f1 and Sox1. (C) Expression of *Pou5f1* and *Sox1* in WT ESCs, *Sall4/1* (S41) dKO ESCs, ESCs overexpressing (OE) Sall4a or Sall4b and their control at day 0, 1, 2 and 3 in N2B27 was measured by qRT-PCR. OE control refers to WT cells expressing the Tet-transactivating factor only. OE control and Sall4 OE cell lines were cultured in doxycycline (DOX) for the entire timecourse. Expression is plotted relative to housekeeping genes as well as to their respective WT controls. Error bars represent s.e.m. between replicates (N=3-5). (D) Alkaline phosphatase (AP) assay of Tet-inducible Sall4 OE cell lines. rtTA refers to the Tet-transactivating factor. Cells were cultured for 96 h in N2B27+DOX (or maintained in 2i/LIF conditions as a control) before replating in 2i/LIF conditions for 5 days. Mean number of AP-positive colonies is shown. Error bars represent s.e.m. (N=4); *P≤0.05, one-way ANOVA followed by Dunnett's multiple comparisons test. Shown below is an example of colonies produced by the indicated ESC lines either with (below) or without (above) 96 h in differentiation conditions. (E) Gene expression analysis across the endoderm differentiation timecourse for the indicated genes in WT, *Sall1* KO, *Sall4* KO and *Sall4/1* dKO cells. Error bars represent s.e.m. between replicates (N=3-9). (F) Gene expression analysis at day 7 of the endoderm differentiation protocol in WT, *Sall1* KO, *Sall4* KO and *Sall4/1* dKO cells. The data are plotted relative to the WT samples. Error bars represent s.e.m. between replicates (N=3-9). Scale bars: 50 µm.

ESC exit from self-renewal and entry into the neural differentiation pathway.

As loss of Sall4 is associated with accelerated differentiation, we predicted that overexpression of Sall4 should result in reduced ESC differentiation. To test this hypothesis, cDNAs encoding Sall4a and Sall4b were expressed either singly or together in a doxycycline-inducible system in WT ESCs (Fig. S1F,G). The Sall4-overexpressing ESCs were then grown in differentiation conditions for 96 h, prior to plating back into 2i/LIF conditions. WT cells expressing the doxycycline-inducible transactivator, but no cDNAs, produced very few alkaline phosphatase-positive colonies after this procedure, indicating that most had undergone lineage commitment (Fig. 2D). By contrast, ESCs overexpressing Sall4 isoforms, either singly or together, produced an increased number of alkaline phosphatase-positive colonies, indicating that overexpression of Sall4 interferes with lineage commitment in ESCs. Further, ESCs overexpressing Sall4 proteins showed persistent *Pou5f1* expression and reduced *Sox1* expression in the neural differentiation timecourse (Fig. 2C; Fig. S1F). Together, these experiments demonstrate that Sall proteins act to slow the pace of neural differentiation in ESC cultures.

To test whether the Sall proteins act as general differentiation inhibitors in ESCs, we next assessed the ability of *Sall4* and *Sall4/1* mutant ESCs to differentiate towards a definitive endoderm fate (Morrison et al., 2008). Although mutant cells were able to silence pluripotency markers and to activate expression of brachyury (*T*), they subsequently failed to activate the endoderm markers *Sox17*, *Foxa2* and *Cxcr4* (Fig. 2E), but neither did they show evidence for having activated a neural programme (Fig. 2F). The failure of *Sall4* null and *Sall4/1* double-null ESCs to adopt either an endodermal or neural fate in this differentiation protocol indicates that Sall4 and Sall1 are not general differentiation inhibitors in ESCs.

Sall4 and Sall1 prevent inappropriate activation of neural genes in ESCs, but are not required for maintenance of the pluripotency GRN

Sall proteins are known to be transcriptional regulators, so we suspected that they would limit neural differentiation by controlling gene expression. To identify the Sall4- and Sall1-dependent transcriptional programmes during ESC self-renewal and during early stages of neural differentiation, we measured global gene expression profiles by RNA-seq in WT, *Sall1* null, *Sall4* null and

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

Sall4/1 double-null ESCs in self-renewing conditions (2i/LIF) and after 48 h in differentiation conditions (N2B27) (Table S1). Global gene expression profiles of WT, *Sall1* null and *Sall4* null ESCs are largely similar in 2i/LIF conditions, resulting in replicates of these genotypes clustering loosely together on the left-hand side of a principal component analysis (PCA) plot (Fig. 3A). By contrast, the double nulls show a distinct profile in the upper middle section of the plot, consistent with increased expression of neural differentiation markers (Fig. S2A) and the presence of morphologically neural cells in 2i/LIF cultures of *Sall4/1* double-null cells (Fig. 1D). After 48 h in differentiation conditions (N2B27) the WT and *Sall1* null ESCs show a similar change in gene expression profiles, moving to the lower right portion of the plot consistent with silencing of pluripotency markers and activation of early differentiation markers (Fig. S2A). *Sall4* null ESCs occupy a somewhat distinct location, presumably owing to partial activation of a neural GRN (Fig. 3A; Fig. S2A). *Sall4/1* double-null cells in N2B27 conditions remain at the top of the plot but move even further to the right, consistent with more complete adoption of a neural phenotype (Fig. S2A).

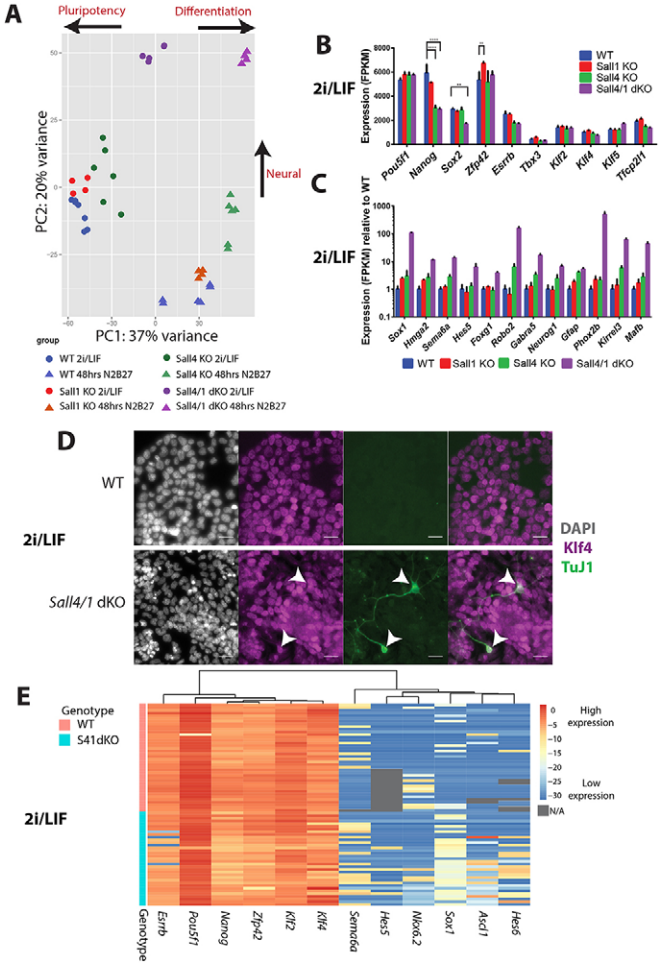


Fig. 3. Sall4 and Sall1 prevent activation of the neurogenesis transcriptional programme but are dispensable for maintaining the pluripotency network. (A) PCA plot representing RNA-seq data from WT, *Sall1* KO, *Sall4* KO and *Sall4/1* dKO cells in self-renewing (2i/LIF) or differentiation (48 h N2B27) culture conditions. Each point represents a separate biological replicate and each genotype is represented by two to three independent cell lines. (B) FPKM values from RNA-seq analysis showing expression of the indicated genes in WT, *Sall1* KO, *Sall4* KO and *Sall4/1* dKO cells in 2i/LIF. Error bars represent s.d. N=4-6 from two to three independent cell lines. $^{**}P \leq 0.01$, $^{****}P \leq 0.0001$, two-way ANOVA followed by a Dunnett's multiple comparison test. (C) Expression of example neural genes is significantly upregulated (see supplementary Materials and Methods) in *Sall4/1* dKO compared with WT cells in 2i/LIF conditions. FPKM values relative to WT levels are shown for all cells. N=4-6 from two to three independent cell lines. Error bars represent s.d. (D) Example immunofluorescence images of WT and *Sall4/1* dKO cells in 2i/LIF stained with DAPI (white) and for Klf4 (magenta) and TuJ1 (green). The right-hand image is a composite of Klf4 and TuJ1. Arrowheads indicate TuJ1-positive cells that co-express Klf4. Scale bars: 25 µm. (E) Heat map constructed from single ESC expression data based on hierarchical clustering for pluripotency-associated genes (*Esrrb*, *Pou5f1*, *Nanog*, *Zfp42*, *Klf2* and *Klf4*) and neural-associated genes (*Sema6a*, *Hes5*, *Nkx6.1*, *Sox1*, *Ascl1* and *Hes6*). Individual cells are ordered from top to bottom: the top 40 are WT cells and the bottom 35 are *Sall4/1* dKO cells. Normalised Ct values (key on the right) refer to –ΔCt values normalised to housekeeping genes (*Atp5a1*, *Ppia* and *Gapdh*). Grey boxes indicate that data are not available (N/A).

The majority of genes found to be misexpressed in either *Sall1* or *Sall4* null ESCs are also misexpressed in *Sall4/1* double-null ESCs, and there is a strong correlation in the direction of the change (Fig. S2B,C). Genes showing increased expression in *Sall4* null or *Sall4/1* null cells show very high enrichment for Gene Ontology (GO) terms involving development, including 'neurogenesis' and 'nervous system development' (Fig. S2D). Further, 42% of genes normally upregulated in WT cells after 48 h in N2B27 are already upregulated in *Sall4/1* double-null cells in 2i/LIF, and the top GO term associated with this group of genes is 'nervous system development' (Fig. S2E). This further supports the hypothesis that Sall4 and Sall1 act together to prevent activation of a neural gene expression programme in ESCs.

Sall4 has been reported to be a component of the pluripotency network, i.e. playing some role in maintaining the GRN underpinning the pluripotent state (Dunn et al., 2014; van den Berg et al., 2010). Findings from the analysis of expression data for individual genes are inconsistent with such a role. Fig. 3B shows that the expression level of many pluripotency-associated genes in ESCs is not significantly altered in the absence of Sall4 and/or Sall1. Although *Sall4* null and *Sall4/1* double-null ESCs show a reduction in *Nanog* expression, and *Sall4/1* double mutants also show a reduction in levels of *Sox2*, this reduction does not result in destabilisation of expression levels of the other pluripotency-associated genes in 2i/LIF.

Although expression of pluripotency markers is largely normal, *Sall4/1* double-null ESC cultures in 2i/LIF conditions expressed elevated levels of genes associated with neuronal differentiation (Table S1; a subset is shown in Fig. 3C). Surprisingly, a fraction of the *Sall4/1* double-null cells in 2i/LIF conditions expressed markers of both a neural (TuJ1) and a pluripotent (Pou5f1 or Klf4) lineage (Fig. 1D, Fig. 3D). In order to expand on this observation we measured gene expression levels in individual ESCs by quantitative RT-PCR (qRT-PCR). As expected, WT ESCs maintained in 2i/LIF conditions robustly expressed pluripotency genes but rarely expressed neural genes (Fig. 3E). *Sall4/1* double-null ESCs showed increased expression of neural genes consistent with RNA-seq and qRT-PCR from bulk cell populations. In addition to aberrant expression of neural genes, individual *Sall4/1* double-null ESCs simultaneously maintained the expression of most pluripotency genes (Fig. 3E). This indicates that components of both the pluripotency and neural differentiation GRNs can be active simultaneously in individual *Sall4/1* double-mutant ESCs. We conclude that in ESCs Sall4 and Sall1 act to prevent activation of neural genes, but are dispensable for maintenance of the pluripotency GRN.

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

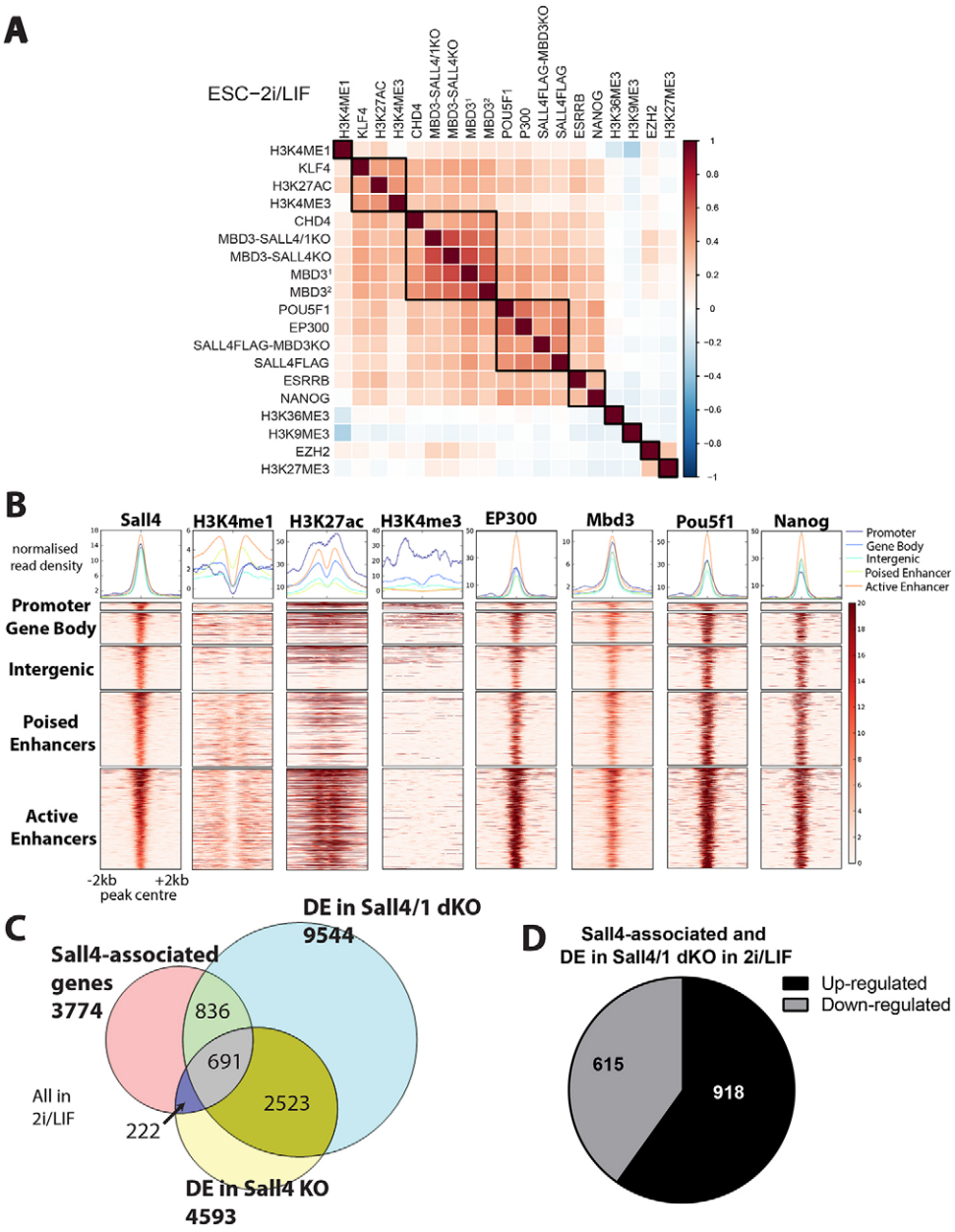


Fig. 4. Sall4 is an enhancer-binding protein and acts to activate as well as inhibit gene expression. (A) Heat map showing correlation between ChIP peaks for the indicated histone modifications and transcription factors in 2i/LIF. ChIP datasets from mutant cells lines are labelled with the antibody used for the ChIP and then the cell line in which they were performed (i.e. MBD3-SALL4KO refers to Mbd3 Chip in *Sall4* KO cells). Mbd3^{1} indicates Mbd3 ChIP performed in the parent line of the *Sall4* KO and Mbd3^{2} indicates Mbd3 ChIP performed in the parent cell line of the *Sall4/1* dKO. Datasets used are listed in the supplementary Materials and Methods. (B) Heat maps of binding profiles of H3K4me1, H3K27ac, H3K4me3, Ep300, Mbd3, Pou5f1 and Nanog shown within 2 kb of the centre of Sall4 peaks. These are partitioned into five groups: promoter-, gene body-, intergenic-, poised enhancer- and active enhancer-associated peaks. Each category is defined in the supplementary Materials and Methods. Graphs above the heat maps show enrichment. (C) Venn diagram showing the overlap between Sall4-associated genes (pink), differentially expressed (DE) genes in *Sall4* KO versus WT (yellow) and differentially expressed genes in *Sall4/1* dKO compared with WT (blue). All in 2i/LIF. (D) Pie chart showing the Sall4-associated genes that are differentially expressed in *Sall4/1* dKO cells compared with WT cells, and the numbers that are upregulated or downregulated. All in 2i/LIF.

Sall4 is an enhancer-binding protein that controls expression of developmental genes

We next sought to identify Sall4-bound genomic sequences in ESCs using ChIP-seq. Previous studies of Sall4 binding to the ESC genome used mouse microarrays (ChIP-Chip), the coverage of which is heavily biased towards genes and promoters, and therefore do not provide genome-wide coverage (Lim et al., 2008; Rao et al., 2010; Tanimura et al., 2013; Yang et al., 2008; Yuri et al., 2009). To facilitate immunoprecipitation of Sall4, the endogenous *Sall4* locus was targeted to add an epitope tag (Avi-3×FLAG) at the C-terminus of the protein (Fig. S3A,B). Immunoprecipitation with an anti-FLAG antibody verified that addition of the epitope tag did not interfere with its known interaction with the NuRD complex (Bode et al., 2016; Kloet et al., 2015; Yuri et al., 2009) (Fig. S3C) nor with its intracellular localisation (Fig. S3D). To verify that addition of the epitope tag did not interfere with normal Sall4 function, ESCs were produced in which both *Sall4* alleles were targeted with the epitope tag. These cells did not show accelerated neural differentiation like *Sall4* null ESCs, and were able to activate endodermal genes when subjected to the endodermal differentiation protocol, unlike *Sall4* null cells (Fig. S3E-G). Thus, addition of a C-terminal epitope tag did not detectably interfere with Sall4 function in ESCs.

Hierarchical clustering of Sall4 ChIP-seq data along with data available in CODEX for a number of transcription factors and histone modifications in ESCs (Sanchez-Castillo et al., 2015) shows that the Sall4 binding profile is well correlated with those of pluripotency-associated transcription factors such as Nanog, Pou5f1, Esrrb and Klf4, as well as for the NuRD component proteins Mbd3 and Chd4 (Fig. 4A,B). Further positive correlation exists with marks of active chromatin H3K4me1, H3K27ac and the histone acetyltransferase Ep300, but not with H3K4me3, consistent with Sall4 associating predominantly with enhancer sequences. Sall4 binding does not correlate with a mark of transcribed gene bodies (H3K36me3), repressive chromatin marks (H3K27me3, H3K9me3) or with a component of the PRC2 complex (Ezh2).

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

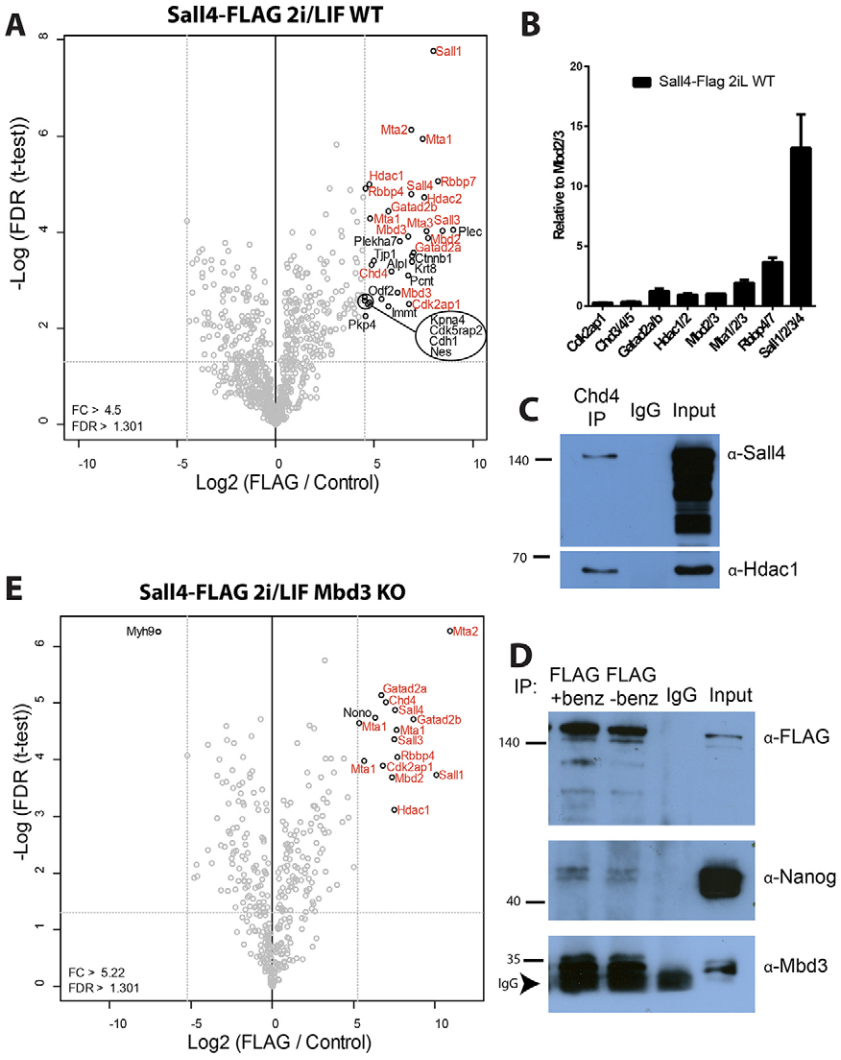


Fig. 5. Identification and stoichiometry of Sall4-interacting proteins in ESCs. (A) Volcano plot showing the significant interactors of Sall4 (black circles) in WT ESCs cultured in 2i/LIF conditions. The proteins highlighted in red are known NuRD components. (B) Stoichiometry of NuRD components and Sall proteins relative to Mbd2/3. Error bars represent s.d. from three independent immunoprecipitations/mass spectrometry replicates. (C) Western blot of immunoprecipitation with anti-Chd4 antibody, IgG control, or 1/10 of input of nuclear extract from WT ESCs and probed with anti-Sall4 (top) or anti-Hdac1 (bottom) antibodies. The anti-Sall4 panel shows a long exposure to visualise the Sall4 band in the Chd4 immunoprecipitation (IP) lane, revealing multiple variously SUMOylated forms of Sall4 in the input lane. Numbers to the left indicate size markers in kDa. (D) Western blot of immunoprecipitation with anti-FLAG antibody, IgG control, or 1/10 of input of nuclear extract from Sall4-FLAG ESCs and probed with the antibodies indicated at right. The anti-FLAG immunoprecipitations are shown with and without the general nuclease benzonase, which makes no difference to the Nanog or Mbd3 association. Size markers are shown on the left in kDa. (E) Volcano plot showing the significant interactors of Sall4 (black circles) in *Mbd3* KO ESCs cultured in 2i/LIF conditions. The proteins highlighted in red are known NuRD components.

To ascertain whether and how the repertoire of Sall4-bound sequences might explain the function of Sall4 in preventing neural differentiation, we assigned each Sall4 peak to its nearest gene. A large proportion of Sall4-associated genes are differentially expressed in the *Sall4/1* double nulls relative to WT cells in either 2i/LIF or N2B27 conditions (40.5% and 43%, respectively; Fig. 4C, Fig. S4A). The genes bound by Sall4 and inappropriately activated in *Sall4/1* double nulls are associated with GO terms involving development and neurogenesis (Fig. S4B), consistent with the crucial function of Sall4 in inhibiting neural specification being to prevent activation of neurogenesis genes in self-renewing conditions and during the early stages of differentiation. Globally, Sall4 is not only a transcriptional repressor, as ∼40% of Sall4-bound and differentially expressed genes show downregulation in the absence of Sall4 (Fig. 4D; Fig. S4C). Notably, several of the GO terms associated with these genes are also associated with genes showing downregulation in WT cells undergoing neural differentiation (Fig. S2E).

NuRD is the major biochemical interactor of Sall4

To better understand how Sall4 exerts its transcriptional regulatory activity, we identified Sall4-interacting proteins in the Sall4-FLAG ESC line using mass spectrometry. As expected, Sall4 robustly co-purified with the core components of the NuRD complex (Fig. 5A, indicated in red). A number of other interacting proteins are shown in Fig. 5A, one of which (Kpna4, an importin subunit) has previously been identified as a NuRD interactor (Kloet et al., 2015). The remainder are proteins normally found in the cytoplasm and/or centriole, whereas in ESCs we find that Sall4 is strictly a nuclear protein (Fig. S3D). Although consistent with the possibility of Sall4 interacting with a centrosome-associated NuRD complex (Sakai et al., 2002; Sillibourne et al., 2007), these were not considered further.

Sall4-FLAG was purified at ∼14-fold excess relative to NuRD [assuming one Mbd2/3 protein and one Sall4 protein per NuRD complex (Kloet et al., 2015)] (Fig. 5B). As we used extraction conditions previously shown to maintain Sall4-NuRD interactions (Kloet et al., 2015), and the Sall4 protein that we purified was expressed from its endogenous locus, this high ratio of Sall4 to NuRD cannot be dismissed as an artefact of protein overexpression or methodology. Immunoprecipitation of Chd4 from WT cells also recovers only a fraction of the Sall4 present in the nucleus, consistent with the majority of Sall4 not being bound to the NuRD complex (Fig. 5C). By contrast, Sall4 immunoprecipitation recovers a large proportion of Mbd3 present in the nucleus, which supports our assertion that the Sall4-NuRD interaction is not being lost due to technical reasons (Fig. 5D). Together, these data show that a relatively minor fraction (∼7%) of Sall4 interacts with the NuRD complex, whereas a large proportion of Mbd3-NuRD contains Sall4.

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

Sall4-FLAG purification was repeated in ESCs lacking Mbd3, a major structural NuRD component protein (Kaji et al., 2006; Reynolds et al., 2012), to identify NuRD-independent interactors of Sall4. Purification of Sall4 and associated proteins in *Mbd3* null ESCs again yielded NuRD components (but no Mbd3), which presumably derive from the small amount of Mbd2-NuRD present in these cells (Fig. 5E). In addition to NuRD components, the only significant interacting protein was the Non-POU domain-containing octamer-binding protein Nono, which was not identified in WT cells. Nono was purified at extremely low levels and is unlikely to be a significant interacting protein.

The pluripotency-associated factors Pou5f1 and Nanog have previously been reported to co-purify with overexpressed Sall4 protein in ESCs grown in serum/LIF conditions (van den Berg et al., 2010), and the endogenous proteins have been shown to interact by immunoprecipitation (Rao et al., 2010; van den Berg et al., 2010; Wu et al., 2006). Although we did identify Pou5f1 peptides in our experiment, these were very few and far below significance (Fig. S5A). Although no Nanog peptides were identified in our mass spectrometry experiments, we were able to detect an interaction between Sall4 and Nanog protein by immunoprecipitation of our tagged Sall4 and western blotting (Fig. 5D). Only a very small proportion of the endogenous nuclear Nanog protein was found to associate with Sall4, which presumably represents a weak and/or infrequent interaction that is below the minimum threshold of detection for mass spectrometry with endogenous Sall4.

We conclude that an interaction between Sall4 protein and pluripotency factors is detectable, and may rise above background in mass spectrometry experiments using overexpressed Sall4 protein, but involves a very small proportion of total endogenous Sall4 protein and thus does not represent a major interaction. Thus, although ∼7% of Sall4 protein is found within the NuRD complex, the majority of Sall4 protein in ESCs does not appear to stably associate with any other protein, but may associate transiently or infrequently with pluripotency-associated transcription factors.

Sall4 neither recruits nor functions through the NuRD complex

Given that a large proportion of NuRD contains Sall4, and that Sall4 and Mbd3 co-occupy a number of genomic sites, the standard model of transcription factor–co-repressor interaction stipulates that Sall4 should recruit NuRD to effect transcriptional repression. If this were true, we would expect that Sall4- and Mbd3-associated genes should show similar changes in expression in *Sall4/1* double-null ESCs as in *Mbd3* null ESCs. Of all Sall4-bound genes showing differential expression in *Sall4/1* double-null ESCs, 20% (315 of 1527 genes) were also bound by Mbd3 and showed transcriptional changes in *Mbd3* null ESCs (Fig. 6A). There is no correlation (neither positive nor negative) in the direction of gene expression changes between *Mbd3* null and *Sall4/1* double-null cells for these 315 genes (Fig. 6B), making it very unlikely that they are co-regulated by Sall4 and NuRD. Similarly, those genes misexpressed in *Sall4/1* or *Mbd3* mutant cells in differentiation conditions (N2B27) show no correlation in terms of the direction of gene expression change (Fig. S6B). Therefore, our analysis provides no evidence that Sall4 and NuRD act in concert to regulate gene expression.

If Sall4 acts to recruit NuRD to specific sites, then we would expect that many Mbd3- and Sall4-bound regions would show loss of Mbd3 binding in *Sall4* null ESCs. Of 4422 Mbd3 peaks lost in *Sall4* null cells, 24% (1073) were bound by Sall4 in WT cells, while 20% of the Mbd3 peaks lost in *Sall4/1* double-null cells were Sall4-bound sites (Fig. 6C; Fig. S6C). This amounts to 7.3% of all Mbd3 sites that could be recruited by Sall4, corresponding to less than 5% of genes misregulated in the *Mbd3* nulls, yet the transcriptional changes seen at these genes in *Mbd3* null ESCs do not correlate with those seen in *Sall4/1* double-null cells (Fig. S6D). If the same analysis is performed using a less stringent method of defining peaks from ChIP replicates (i.e. by merging replicates rather than using the IDR method; see Materials and Methods), then 3.0% of Mbd3 peaks show both Sall4 dependency and Sall4 binding (Fig. S6E). Thus, we find no evidence to support a model whereby Sall4 directs the recruitment of NuRD to control gene expression in ESCs.

Sall4 occupies enhancers with pluripotency-associated transcription factors to regulate transcription

Although Sall4 does not dictate NuRD chromatin targets, surprisingly, NuRD was found to influence the genome-wide distribution of Sall4. ChIP-seq for Sall4-FLAG in *Mbd3* null cells identified 3.5-fold more Sall4-bound locations than in WT cells (17,739 versus 5062; Fig. 6D). The Sall4-bound sites found only in *Mbd3* null cells predominantly consisted of enhancers, as is seen for the WT cohort of Sall4-bound sites (Fig. S6F). In addition to Sall4 binding to novel sites in the absence of Mbd3/NuRD (e.g. *Tex13* and *Ppp2r2c* enhancers; Fig. S7A,B), Sall4 also shows increased binding at some peaks seen in WT cells (e.g. *Nanog*, but not *Pou5f1*; Fig. S7A,B). This indicates that more Sall4 protein is available to bind chromatin in the absence of Mbd3. Indeed, *Mbd3* null ESCs contain moderately (2- to 3-fold) increased levels of Sall4 protein, despite there being no increase in *Sall4* transcript levels (Table S1, Fig. S5B).

By focusing on the Sall4-enriched regions seen only in *Mbd3* null ESCs, we were able to investigate the consequences of novel Sall4 binding to enhancer sequences. In *Mbd3* null cells, they not only gain Sall4 protein enrichment but also become enriched for the pluripotency-associated transcription factors Pou5f1, Nanog, Klf4 and Esrrb (Fig. 6E). Notably, no increase in Chd4 protein enrichment is seen at these same sites in the *Mbd3* null cells, indicating that the observed increase in transcription factor association is not simply a consequence of these sites becoming generally more accessible. What consequence does recruitment of transcription factors have on these enhancers? Assigning these sites to their nearest genes identifies 6666 genes, of which nearly one-fifth (1166) show a significant gene expression change in *Mbd3* null ESCs (Fig. S7C), with approximately equal numbers showing increased or decreased expression (Fig. S7D). These sites are not associated with significant Mbd3 enrichment in WT cells (Fig. 6D), yet they account for nearly one-third (1166/4049) of all genes misexpressed in *Mbd3* null ESCs (Fig. S7C).

Recruitment of Sall4 and four pluripotency-associated transcription factors to these enhancers is equally likely to result in gene activation as it is in repression. GO terms associated with upregulated genes involve development and motility (Fig. S7E), whereas genes showing decreased expression do not significantly associate with any specific GO term. Thus, enhancers able to increase transcription in response to the recruitment of this group of transcription factors are predominantly associated with developmental genes, whereas enhancers associated with other kinds of genes are not activated by these transcription factors, and indeed this recruitment interferes with transcription. In summary, we propose that Sall4 acts to prevent neural differentiation of ESCs by binding to enhancers along with other pluripotency-associated transcription factors, where their presence interferes with gene

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

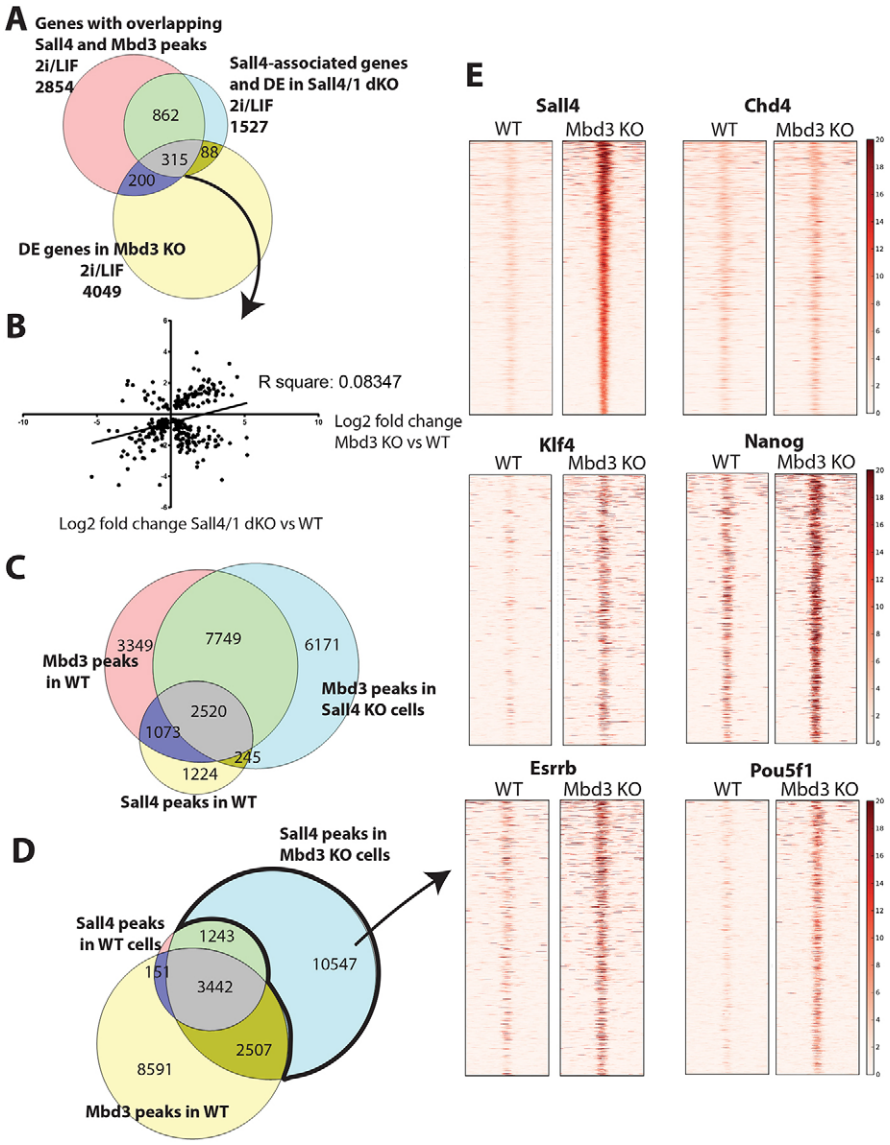


Fig. 6. Sall4 neither recruits nor acts through the NuRD complex. (A) Venn diagram showing the overlap between genes associated with overlapping Sall4 and Mbd3 peaks in 2i/LIF (pink), genes associated with a Sall4 peak and differentially expressed in *Sall4/1* dKO in 2i/LIF (blue), and genes differentially expressed in *Mbd3* KO cells in 2i/LIF (yellow). (B) Plot comparing the $\log_2$ fold change between differentially expressed genes in both *Sall4/1* dKO and *Mbd3* KO cells compared with WT cells. These genes have Sall4 and Mbd3 overlapping peaks and are differentially expressed in *Sall4/1* dKO cells as well as in *Mbd3* KO cells compared with WT, all in 2i/LIF (i.e. the 315 genes shown in grey in A). A linear regression was performed to generate the R-square value. (C) Venn diagram showing overlap between Mbd3 peaks in WT cells (pink), Mbd3 peaks in *Sall4* KO cells (blue) and Sall4 peaks in WT cells (yellow). The WT used for this comparison is the parent line of the *Sall4* KO cells. All are in 2i/LIF. (D) Venn diagram showing the overlap of Sall4 peaks in WT cells (pink) with Sall4 peaks in *Mbd3* KO cells (blue) and Mbd3 peaks in WT cells (yellow). (E) ChIP-seq heat maps of 2 kb either side of the Sall4 peaks only found in *Mbd3* KO ESCs and not normally bound by Mbd3 (bold black outline in D) for the indicated transcription factors in WT and *Mbd3* KO ESCs.

activation (Fig. 7). At other enhancers, binding of this same cohort of transcription factors increases transcription.

DISCUSSION

Sall4 is an essential protein for early mammalian development. Here, we show that Sall4 and Sall1 function to prevent activation of neural development genes in ESCs, but are not required to maintain the pluripotent state. Sall4 is predominantly an enhancer-binding protein and, although it binds to a similar array of genomic locations as the pluripotency-associated proteins Pou5f1, Nanog, Esrrb and Klf4, it does not stably associate with these proteins. We further clarify the nature of the relationship between Sall4 and the NuRD complex. Although a proportion of Sall4 protein does stably interact with the NuRD complex, contrary to the standard model of co-repressor recruitment to DNA, Sall4 neither recruits NuRD to specific sites on DNA nor does it use NuRD to control expression of its target genes. Rather, Sall4 occupancy of enhancer sequences, along with other pluripotency-associated transcription factors, can either enhance or interfere with transcription, depending upon the target gene (Fig. 7).

Previous studies of Sall4 function in ESCs have produced conflicting conclusions about the role of Sall4 in ESC self-renewal (Rao et al., 2010; Sakaki-Yumoto et al., 2006; Tsubooka et al., 2009; Yuri et al., 2009; Zhang et al., 2006). It is very likely that the differing results obtained from these various laboratories are heavily influenced by the different culture conditions. Using a fully defined culture system [2i/LIF (Ying et al., 2008)] we show that Sall4 and Sall1 are dispensable for ESC self-renewal, but that they prevent premature activation of neural genes. This result agrees with a report that knockdown of Sall4 in 2i/LIF conditions does not significantly compromise ESC self-renewal (Dunn et al., 2014), and with our finding that Sall4 overexpression in WT cells inhibits neural differentiation (Fig. 2C). Notably, Yuri et al. (2009) were able to establish Sall4/1-double-knockout ESCs in serum/LIF conditions, indicating that this is not a difference in Sall4 function between different culture conditions. We show clearly that Sall4 is not an essential pluripotency factor, but rather is a differentiation inhibitor. We speculate that interference with Sall protein activity might enhance the efficiency of directed pluripotent cell neural differentiation protocols for disease modelling or regenerative medicine applications.

Using mass spectrometry on immunoprecipitated endogenous Sall4 protein, we find that ~7% of nuclear Sall4 protein interacts with the NuRD complex. The simplest interpretation of this would

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

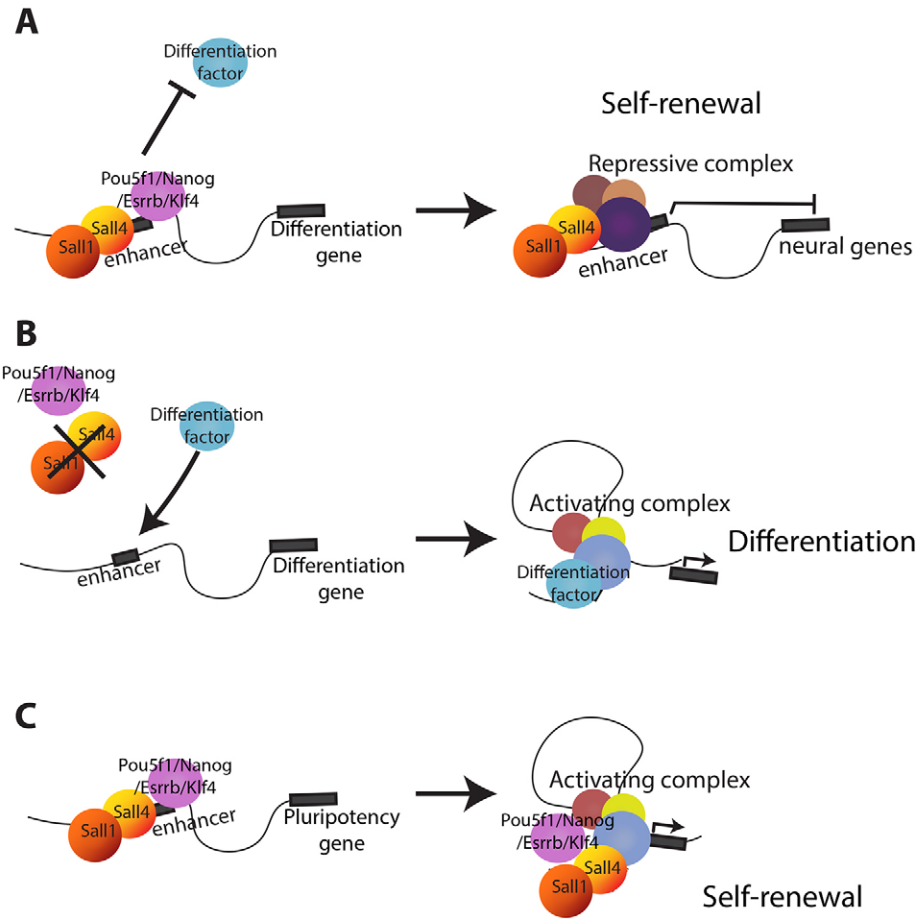


Fig. 7. Model of Sall4 activity in ESCs. (A) Sall4 binds to the enhancer of a gene normally expressed during neural development along with Nanog, Pou5f1, Klf4 and Esrrb, preventing the association of a lineage-specific transcription factor (differentiation factor). The result is failure to activate the neural gene. (B) In the absence of Sall4 and Sall1 the differentiation factor is no longer prevented from binding to the enhancer and the neural gene is then inappropriately activated. (C) At enhancers of genes normally expressed in undifferentiated ESCs the binding of Sall proteins along with Pou5f1, Nanog, Klf4 and Esrrb maintains transcriptional activation of the self-renewal gene.

be that this subset of Sall4 indirectly influences gene expression by recruiting NuRD to specific sequences. The problem with this scenario is that Sall4 and NuRD serve opposing functions in ESCs: NuRD facilitates exit of ESCs from the self-renewing state by restricting expression levels of pluripotency-associated genes (Reynolds et al., 2012), whereas Sall4 acts to prevent activation of neural genes and precocious neural specification in ESCs (Fig. 2). This is not what one would expect if NuRD collaborates with Sall4 to regulate the expression of Sall4 target genes. We find no evidence that Sall4 plays any significant role in recruiting NuRD to chromatin, nor that the expression levels of Sall4 target genes are sensitive to the presence or absence of NuRD.

We identified neither Nanog nor Pou5f1 as a significant Sall4-interacting protein in our proteomics experiments, although a weak interaction could be detected by immunoprecipitation and western blotting (Fig. 5D). Both of these proteins have been identified as Sall4 interactors in other studies (Rao et al., 2010; van den Berg et al., 2010; Wu et al., 2006). Our study differs from previous studies of Sall4 interactors as we have incorporated an epitope tag to the endogenous *Sall4* locus, and therefore have not introduced an extra copy of *Sall4* into ESCs. Sall4 dosage is important in somatic tissues (Koshiba-Takeuchi et al., 2006; Sakaki-Yumoto et al., 2006), so introduction of more Sall4, even if expressed at levels comparable to endogenous protein, would increase the concentration of nuclear Sall4 and might enable association with proteins such as Pou5f1 and Nanog (Fig. 7). Our results do not preclude an interaction between Sall4 and these pluripotency factors, but rather suggest that such interactions are either transient or involve only a minor fraction of total Sall4.

The Sall4 protein does not appear to have any enzymatic activity, does not recruit the NuRD complex to its sites of action, and does not have any other major, stably interacting proteins. Sall4 is predominantly found at enhancers, which are also often bound by pluripotency-associated transcription factors such as Pou5f1, Nanog, Klf4 and Esrrb (Fig. 4A,B, Fig. 7). Loss of Sall4 results in increased transcription of some Sall4-associated genes and reduced transcription of others, indicating that the outcome of Sall4 activity depends upon the sequence to which it binds (Fig. 7). We propose that accumulation of these transcription factors at enhancers that normally respond to lineage-specific transcription factors interferes with their activation, possibly by steric hindrance of transcription factor binding (Fig. 7A). In cells lacking the Sall proteins, this accumulation of transcription factors at neural genes does not occur and permits gene activation (Fig. 7B). By contrast, binding of these proteins to enhancers of genes normally expressed during ESC self-renewal promotes or enforces active transcription (Fig. 7C), although maintaining expression of pluripotency-associated genes does not strictly require the presence of Sall4. This scenario is similar to that seen for Pou5f1 during reprogramming, where Pou5f1 binding to enhancers of somatic genes in mouse embryonic fibroblasts correlates with transcriptional silencing of the associated gene, whereas Pou5f1 binding to enhancers of genes normally expressed in pluripotent cells correlates with activation (Chen et al., 2016). This would also mean that the dosage of Sall4 would be very important: too little Sall4 and some genes might be activated inappropriately, while too much Sall4 could interfere with the expression of lineage-appropriate genes. This could explain the observed haploinsufficiency of Sall4 during mammalian

Development (2016) 143, 3074-3084 doi:10.1242/dev.139113

development (Koshiba-Takeuchi et al., 2006; Sakaki-Yumoto et al., 2006).

MATERIALS AND METHODS

Mouse ESC lines, culture and manipulation

All ESC lines were cultured in 2i/LIF conditions on gelatin-coated plates. ESC derivations were performed in 2i/LIF conditions. Gene targeting was carried out using homologous recombination methods and verified by long-range PCR, RT-PCR and western blotting. For details, including the antibodies used, see the supplementary Materials and Methods. Doxycycline treatment with alkaline phosphatase staining, the neural differentiation protocol and teratoma assay were performed as detailed in the supplementary Materials and Methods. All animal experiments were approved by the Animal Welfare and Ethical Review Body of the University of Cambridge and carried out under appropriate UK Home Office licenses.

Chromatin immunoprecipitation and sequencing

ChIP-seq in ESCs was performed as previously described (Reynolds et al., 2012). For details, including the antibodies used, see the supplementary Materials and Methods.

Bioinformatic analyses

Sall4-FLAG, Mbd3 and Chd4 ChIP-seq data were analysed using the irreproducible discovery rate (IDR) method, which assesses replicate agreement and therefore only calls peaks that are strong in all replicates (Landt et al., 2012; Li et al., 2011). This has the effect of removing false positives, but also of removing many weaker true positives. Thus, the set of 'bound' peaks used in the subsequent analyses is not comprehensive, but is of very high confidence and will represent only the strongest-bound peaks. Differentially expressed genes are listed in Table S1, and genes closest to Sall4 and Mbd3 peaks are listed in Table S2. For full details, see the supplementary Materials and Methods.

Mass spectrometry

To identify Sall4 interactors, tryptic peptides obtained from affinity-purified nuclear proteins were subject to mass spectrometry analysis and LFQ peptide identification as described in the supplementary Materials and Methods.

qRT-PCR

Single-cell expression analysis of pluripotency and lineage markers was performed by qRT-PCR using the TaqMan primers described in the supplementary Materials and Methods.

Acknowledgements

We thank Bill Mansfield, Charles Etienne-Dumeau, Maike Paramour, Peter Humphreys and SCI Tissue Culture Staff for excellent technical assistance; members of the B.H. lab for stimulating discussions; and Austin Smith, Aoife O'Shaughnessy-Kirwan, Peter Kirwan and B.H. lab members for critical comments on the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

A.M., S.L.K. and B.H. performed the experiments; A.M., M.R., S.L.K. and R.L. analysed data; R.N. provided reagents; M.V., P.B. and B.H. supervised projects; and A.M. and B.H. wrote the paper.

Funding

A.M. was supported by a Wellcome Trust Four Year PhD Studentship. Work in the B.H. lab is supported by a Wellcome Trust Senior Fellowship in the Basic Biomedical Sciences [098021/Z/11/Z], and through Wellcome Trust and UK Medical Research Council core funding to the Cambridge Stem Cell Institute [079249/Z/06/I]. The B.H. and M.V. labs were further supported by the European Union Seventh Framework Programme (FP7) Project '4DCellFate'. Deposited in PMC for immediate release.

Data availability

ChIP-seq and RNA-seq data have been deposited in ArrayExpress: ChIP-seq data: E-MTAB-4565; RNA-seq data: E-MTAB-4566. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org/) via the PRIDE partner repository with the dataset identifier PXD003614.

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.139113.supplemental

References

Al-Baradie, R., Yamada, K., St Hilaire, C., Chan, W.-M., Andrews, C., McIntosh, N., Nakano, M., Martonyi, E. J., Raymond, W. R., Okumura, S. et al. (2002). Duane radial ray syndrome (Okihiro syndrome) maps to 20q13 and results from mutations in SALL4, a new member of the SAL family. *Am. J. Hum. Genet.* **71**, 1195-1199.

Allen, H. F., Wade, P. A. and Kutateladze, T. G. (2013). The NuRD architecture. *Cell. Mol. Life Sci.* **70**, 3513-3524.

Bode, D., Yu, L., Tate, P., Pardo, M. and Choudhary, J. (2016). Characterization of two distinct Nucleosome Remodeling and Deacetylase (NuRD) complex assemblies in embryonic stem cells. *Mol. Cell. Proteomics* **15**, 878-891.

Chen, J., Chen, X., Li, M., Liu, X., Gao, Y., Kou, X., Zhao, Y., Zheng, W., Zhang, X., Huo, Y. et al. (2016). Hierarchical Oct4 binding in concert with primed epigenetic rearrangements during Somatic Cell Reprogramming. *Cell Rep.* **14**, 1540-1554.

de Celis, J. F. and Barrio, R. (2009). Regulation and function of Spalt proteins during animal development. *Int. J. Dev. Biol.* **53**, 1385-1398.

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

Elling, U., Klasen, C., Eisenberger, T., Anlag, K. and Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. *Proc. Natl. Acad. Sci. USA* **103**, 16319-16324.

Kaji, K., Caballero, I. M., MacLeod, R., Nichols, J., Wilson, V. A. and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat. Cell Biol.* **8**, 285-292.

Kanda, S., Tanigawa, S., Ohmori, T., Taguchi, A., Kudo, K., Suzuki, Y., Sato, Y., Hino, S., Sander, M., Perantoni, A. O. et al. (2014). Sall1 maintains nephron progenitors and nascent nephrons by acting as both an activator and a repressor. *J. Am. Soc. Nephrol.* **25**, 2584-2595.

Kiefer, S. M., McDill, B. W., Yang, J. and Rauchman, M. (2002). Murine Sall1 represses transcription by recruiting a histone deacetylase complex. *J. Biol. Chem.* **277**, 14869-14876.

Kloet, S. L., Baymaz, H. I., Makowski, M., Groenewold, V., Jansen, P. W. T. C., Berendsen, M., Niazi, H., Kops, G. J. and Vermeulen, M. (2015). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS J.* **282**, 1774-1785.

Kohlhase, J., Wischermann, A., Reichenbach, H., Froster, U. and Engel, W. (1998). Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome. *Nat. Genet.* **18**, 81-83.

Kohlhase, J., Heinrich, M., Schubert, L., Liebers, M., Kispert, A., Laccone, F., Turnpenny, P., Winter, R. M. and Reardon, W. (2002). Okihiro syndrome is caused by SALL4 mutations. *Hum. Mol. Genet.* **11**, 2979-2987.

Koshiba-Takeuchi, K., Takeuchi, J. K., Arruda, E. P., Kathiriya, I. S., Mo, R., Hui, C.-C., Srivastava, D. and Bruneau, B. G. (2006). Cooperative and antagonistic interactions between Sall4 and Tbx5 pattern the mouse limb and heart. *Nat. Genet.* **38**, 175-183.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P. et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813-1831.

Lauberth, S. M. and Rauchman, M. (2006). A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J. Biol. Chem.* **281**, 23922-23931.

Le Guezennec, X., Vermeulen, M., Brinkman, A. B., Hoeijmakers, W. A. M., Cohen, A., Lasonder, E. and Stunnenberg, H. G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* **26**, 843-851.

Li, Q., Brown, J. B., Huang, H. and Bickel, P. J. (2011). Measuring reproducibility of High-Throughput Experiments. *Ann. Appl. Stat.* **5**, 1752-1779.

Lim, C. Y., Tam, W.-L., Zhang, J., Ang, H. S., Jia, H., Lipovich, L., Ng, H.-H., Wei, C.-L., Sung, W. K., Robson, P. et al. (2008). Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. *Cell Stem Cell* **3**, 543-554.

Lu, J., Jeong, H. W., Kong, N., Yang, Y., Carroll, J., Luo, H. R., Silberstein, L. E., Yupoma, and Chai, L. (2009). Stem cell factor SALL4 represses the

transcriptions of PTEN and SALL1 through an epigenetic repressor complex. *PLoS ONE* **4**, e5577.
Morey, L., Santanach, A. and Di Croce, L. (2015). Pluripotency and epigenetic factors in mouse embryonic stem cell fate regulation. *Mol. Cell. Biol.* **35**, 2716-2728.
Morrison, G. M., Oikonomopoulou, I., Migueles, R. P., Soneji, S., Livigni, A., Enver, T. and Brickman, J. M. (2008). Anterior definitive endoderm from ESCs reveals a role for FGF signaling. *Cell Stem Cell* **3**, 402-415.
Nishinakamura, R., Matsumoto, Y., Nakao, K., Nakamura, K., Sato, A., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Scully, S., Lacey, D. L. et al. (2001). Murine homolog of SALL1 is essential for ureteric bud invasion in kidney development. *Development* **128**, 3105-3115.
Niwa, H. (2007). How is pluripotency determined and maintained? *Development* **134**, 635-646.
Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M. M. and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* **6**, 382-395.
Rao, S., Zhen, S., Roumiantsev, S., McDonald, L. T., Yuan, G.-C. and Orkin, S. H. (2010). Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. *Mol. Cell. Biol.* **30**, 5364-5380.
Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T. et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* **10**, 583-594.
Sakai, H., Urano, T., Ookata, K., Kim, M.-H., Hirai, Y., Saito, M., Nojima, Y. and Ishikawa, F. (2002). MBD3 and HDAC1, two components of the NuRD complex, are localized at Aurora-A-positive centrosomes in M phase. *J. Biol. Chem.* **277**, 48714-48723.
Sakaki-Yumoto, M., Kobayashi, C., Sato, A., Fujimura, S., Matsumoto, Y., Takasato, M., Kodama, T., Aburatani, H., Asashima, M., Yoshida, N. et al. (2006). The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. *Development* **133**, 3005-3013.
Sanchez-Castillo, M., Ruau, D., Wilkinson, A. C., Ng, F. S. L., Hannah, R., Diamanti, E., Lombard, P., Wilson, N. K. and Gottgens, B. (2015). CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res.* **43**, D1117-D1123.
Signolet, J. and Hendrich, B. (2015). The function of chromatin modifiers in lineage commitment and cell fate specification. *FEBS J.* **282**, 1692-1702.
Sillibourne, J. E., Delaval, B., Redick, S., Sinha, M. and Doxsey, S. J. (2007). Chromatin remodeling proteins interact with pericentrin to regulate centrosome integrity. *Mol. Biol. Cell* **18**, 3667-3680.
Sweetman, D. and Munsterberg, A. (2006). The vertebrate spalt genes in development and disease. *Dev. Biol.* **293**, 285-293.
Tanimura, N., Saito, M., Ebisuya, M., Nishida, E. and Ishikawa, F. (2013). Stemness-related factor Sall4 interacts with transcription factors Oct-3/4 and Sox2 and occupies Oct-Sox elements in mouse embryonic stem cells. *J. Biol. Chem.* **288**, 5027-5038.
Tsubooka, N., Ichisaka, T., Okita, K., Takahashi, K., Nakagawa, M. and Yamanaka, S. (2009). Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. *Genes Cells* **14**, 683-694.
van den Berg, D. L. C., Snoek, T., Mullin, N. P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I. and Poot, R. A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**, 369-381.
Warren, M., Wang, W., Spiden, S., Chen-Murchie, D., Tannahill, D., Steel, K. P. and Bradley, A. (2007). A Sall4 mutant mouse model useful for studying the role of Sall4 in early embryonic development and organogenesis. *Genesis* **45**, 51-58.
Wu, Q., Chen, X., Zhang, J., Loh, Y.-H., Low, T.-Y., Zhang, W., Zhang, W., Sze, S.-K., Lim, B. and Ng, H.-H. (2006). Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J. Biol. Chem.* **281**, 24090-24094.
Xiong, J. (2014). SALL4: engine of cell stemness. *Curr. Gene Ther.* **14**, 400-411.
Yang, J., Chai, L., Fowles, T. C., Alipio, Z., Xu, D., Fink, L. M., Ward, D. C. and Ma, Y. (2008). Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **105**, 19756-19761.
Yang, J., Gao, C., Chai, L. and Ma, Y. (2010). A novel SALL4/OCT4 transcriptional feedback network for pluripotency of embryonic stem cells. *PLoS ONE* **5**, e10766.
Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. and Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183-186.
Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523.
Yuri, S., Fujimura, S., Nimura, K., Takeda, N., Toyooka, Y., Fujimura, Y.-I., Aburatani, H., Ura, K., Koseki, H., Niwa, H. et al. (2009). Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression. *Stem Cells* **27**, 796-805.
Zhang, J., Tam, W.-L., Tong, G. Q., Wu, Q., Chan, H.-Y., Soh, B.-S., Lou, Y., Yang, J., Ma, Y., Chai, L. et al. (2006). Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat. Cell Biol.* **8**, 1114-1123.
Zhang, X., Yuan, X., Zhu, W., Qian, H. and Xu, W. (2015). SALL4: an emerging cancer biomarker and target. *Cancer Lett.* **357**, 55-62.

Supplementary Information

Supplementary Methods

ES cell lines, culture and manipulation

Mouse ES cells were grown on 0.1% gelatin in 2i/LIF conditions (Ying et al., 2008). ES cell derivation was performed in 2i conditions as described (Nichols et al., 2009). Sall4 conditional (Flox/Flox) ES cell lines were made via two different procedures. In the first case an *Mbd3*$^{+/-}$ ES cell line was subjected to two rounds of traditional gene targeting with a *Sall4* conditional targeting construct (Sakaki-Yumoto et al., 2006) to generate a homozygous floxed ES cell line (2 independently-derived replicate lines). In the second ES cells were derived from *Sall4*$^{Flox/Flox}$ mice through homozygous intercrosses. *Sall4*$^{Flox/Flox}$*Sall1*$^{Flox/Flox}$ ES cells were derived from Sall4/Sall1 compound homozygous floxed mice (Yuri et al., 2009). Sall4Flox/Flox and *Sall4*$^{Flox/Flox}$*Sall1*$^{Flox/Flox}$ ES cells were transfected with a Mer-Cre-Mer expression construct (kindly provided by Elly Tanaka) to enable inducible deletion of Floxed alleles. No recombination between LoxP sites was detected in the absence of induction for any of the ES cell lines. Sall4 null lines and Sall4/1 double null lines were made through induction of Cre and selection for clones showing homozygous deletion (3 and 2 biological replicates, respectively). Sall1 null lines (*Sall4*$^{Flox/\Delta}$*Sall1*$^{\Delta/\Delta}$) were made by Cre induction in the double Floxed ES cell line and selection for *Sall1* deletion (two biological replicates). Sall4-FLAG cell lines were made in an *Mbd3*$^{Flox/-}$ ES cell line (Kaji et al., 2006) by traditional gene targeting. One heterozygous (*Sall4*$^{FLAG/+}$) and one homozygous (*Sall4*$^{FLAG/FLAG}$) ES cell line was made. Cre transfection was used to create the *Mbd3*$^{-/-}$ *Sall4*$^{Flox/+}$ ES cell line. All phase contrast images were captured using Olympus IX51 microscope. Gene targeting was carried out using traditional homologous recombination methods and verified by long-range PCR,

cDNAs encoding Sall4 isoforms were cloned into a Tet-ON piggy-BAC plasmid (kindly provided by Jose Silva) and transfected along with the doxycycline-controlled transactivator (rtTA) into E14tg2a cells maintained in 2i/LIF conditions. 10^5 cells were then transferred to N2B27 media including 1µg/ml doxycycline for 96 hours. Subsequently, 600 cells/well were plated into a lamin-coated 6-well dish in 2i/LIF and allowed to grow for 5 days prior to fixing and staining. Alkaline phosphatase staining was performed (Sigma 86R-1KT) following the manufacturer's instructions.

Neural differentiation: ES cells were plated onto Laminin coated plates in N2B27 media at a density of 100,000 per 6-well. The media was replaced every day and the cells were not passaged during this timecourse. The cells were harvested at different time points for RNA or for immunofluorescence. All fluorescent images were captured on either Zeiss Apotome or Leica DMI3000 microscopes. Image J was used for all image preparations.

Teratoma assays: ES cells grown in 2i/LIF were harvested and resuspended into N2B27 media. These were placed on ice and the SCI transgenic facility injected 15µl to 20µl of the cells into the kidney capsule of NOD-SCID mice. After 2 to 3 months the mice were sacrificed. The teratoma was placed in 10% formal saline overnight, followed by 70% ethanol before being sectioned for H&E staining and for parafin sections by SCI Histology facility. All animal experiments were approved by the Animal Welfare and Ethical Review Body of the University of Cambridge and carried out under appropriate UK Home Office licenses.

Chromatin immunoprecipitation and sequencing

Chromatin immunoprecipitation was performed as described (Reynolds et al., 2012). Library preparation for sequencing was performed according to standard methods. First RiboZero (depletion of ribosomal RNA), then 60-100ng of depleted RNA was used as the input into the Bioo Scientific NextFlex rapid directional RNA kit (#5138-08). ChIP libraries were made using the Bioo Scientific NEXTflex Rapid DNA-Seq Kit (5144-02). Size selection was performed on the pool if required to ensure small DNA size fragments were sequenced.

Bioinformatic analyses

Sall4-FLAG, Mbd3 and Chd4 ChIP-seq data were analysed using the irreproducible discovery rate (IDR) method which assesses replicate agreement and therefore only calls peaks that are strong in all replicates (Landt et al., 2012; Li et al., 2011). This has the effect of removing false positives, but also of removing many weaker true positives. Therefore the set of 'bound' peaks used in the subsequent analyses is not comprehensive, but is of very high confidence and will represent only the strongest-bound peaks.

ChIP Seq data were aligned to the mm10 reference genome using bowtie version 0.12.8. Only uniquely mapped reads were used for further analysis. Irreproducible Discovery Rate (IDR) framework in combination with the MACS2 peak caller version 2.1.0 was used to identify reproducible peaks for each condition. Peak calling was first done using MACS2 peak caller (Liu, 2014) at a p value threshold of 0.001, using input as control, to allow for

sufficient false peaks for the Irreproducible Discovery Rate (IDR) between replicates. All calculations were carried out on peaks identified by irreproducible discovery rate (IDR) analysis with a cutoff of 0.05 (Landt et al., 2012).

Promoters were defined as -2000bp and +500bp around Ensembl (Ensembl Release 75) annotated transcription start sites. Peaks overlapping by at least 1bp with a promoter region were considered as promoter peaks. Active enhancers were called based on overlap with H3K27ac, H3K4me1 and no overlap with H3K4me3, and poised enhancers were marked by H3K4me1, no H3K4me3, and excluding the active enhancers. The peaks overlapping at least 1bp with a gene body region were classified as gene body peaks. Other peaks were called intergenic.

The Ezh2, H3K27me3, H3K36me3, H3K4me3, H3K9me3 ChIP-seq datasets are from (Marks et al., 2012) and the H3K27ac, H3K4me1, EP300 ChIP-seq datasets are from (Buecker et al., 2014). The Oct4 ChIP-seq dataset used was a combination of (Buecker et al., 2014) and our own.

Paired-end RNA-seq reads were aligned to the reference genome GRCm38/mm10, downloaded from the Ensembl database ftp.ensembl.org using Gsnap version gmap-2014-12-17. Only uniquely mapped reads were used for further analysis. Gene counts from SAM files were obtained using htseq-count version 0.6.1 with mode intersection non-empty, -s reverse. The gene annotation was extracted from Ensembl Gene Release 75. Differential gene expression analysis was conducted using Bioconductor R (R-3.1.2) package DESeq2 version 1.6.3. An adjusted p Value threshold of 0.05 was used to determine differential gene expression. Expression values are FPKM (read number normalized across samples by DESeq size factor and across genes by gene length) calculated using custom R-scripts. ChIP-seq and RNAseq data have been deposited in ArrayExpress: ChIP-seq data: E-MTAB-4565, RNA-seq data: E-MTAB-4566.

GO analyses were performed using HumanMine: http://www.humanmine.org/, and Venn diagrams were made using http://omics.pnl.gov. Statistics were performed using GraphPad Prism. Datasets used include those in Tables S1, S2 and data in CODEX (Sanchez-Castillo et al., 2015).

FLAG pulldowns and label-free quantitation (LFQ) LC-MS/MS analysis

Label-free FLAG pulldowns were performed in triplicate as previously described (Smits et al., 2013) with the following modifications. For each pulldown, 2 mg of nuclear extract

was incubated with 10 µl anti-FLAG M2 affinity gel (Sigma) and 50 µg/mL ethidium bromide in a total volume of 400 µl. Affinity purified proteins were subject to on-bead trypsin digestion as previously described (Baymaz et al., 2014). Tryptic peptides were acidified and desalted using StageTips (Rappsilber et al., 2007) prior to mass spec analyses. Tryptic peptides were separated with an Easy-nLC 1000 (Thermo Scientific). Buffer A was 0.1% formic acid and Buffer B was 80% acetonitrile and 0.1% formic acid. Peptides were separated using a 94-min gradient from 9-32% Buffer B followed by washes at 50% then 95% Buffer B for 120 min of total data collection time. Mass spectra were recorded on an LTQ-Orbitrap QExactive mass spectrometer (Thermo Fisher Scientific), selecting the top 10 most intense precursor ions for fragmentation.

LFQ peptide identification and analysis

Thermo RAW files from LFQ AP-MS/MS were analysed with MaxQuant version 1.5.1.0 using default settings and searching against the UniProt curated mouse proteome, release 2014_01. Additional options for Match between runs, LFQ, and iBAQ were selected. Stoichiometry calculations and volcano plots were produced essentially as described (Smits et al., 2013) using Perseus version 1.4.0.8 and in-house R scripts. Statistical cutoffs were chosen such that no proteins were present as outliers on the control, non-FLAG side of the volcano plot. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD003614.

Antibodies used in this work

Primary Antibody	Dilution for Immunofluorescence	Dilution for Western Blotting	Amount used for ChIP	Company, Cat No.
α-Sall4	-	1:1000	-	Abcam ab-29112-100
α -Sall4	1:200	1:1000	-	CosmoBio PPX-PP-PPZ0601-00
α -FLAG	1:200	1:2000	3µg	(M2) Sigma F3165
α -Mbd3	-	1:1000	15µg	Bethyl A302528A
α -Chd4	-	1:5000	10µg	Abcam ab70469
α -Esrrb	-	-	2.5µg	R&D Systems PP-H6705-00
α -Oct4	1:200	-	5µg	Santa Cruz Biotechnology sc-8628
α -Nanog	-	-	2.5µg	Bethyl A300-397A
α -Klf4	1:200	-	2.5µg	R&D Systems AF3158

α -Tuj1	1:500	-	-	Covance, MMS-435P
α -FoxA2	1:200	-	-	Santa Cruz sc-6554
α -Sox1	1:200	-	-	Chemicon, AB5768
α Mta1/2	-	1:1000	-	Santa Cruz, sc-9447
α -Hdac1	-	1:2000	-	Abcam, 7028
α – RNAPolII S5P (CTD4H8)		1:5000		Millipore, 05-623

ChIP-seq datasets produced in this study:

Dataset	Cell line	ChIP	Replicates
2lox_2i-Esrrb	$Mbd3^{Flox/-}$	Esrrb	1
2lox_2i-Klf4	$Mbd3^{Flox/-}$	Klf4	1
2lox_2i-Nanog	$Mbd3^{Flox/-}$	Nanog	1
2lox_2i-Oct4	$Mbd3^{Flox/\Delta}$	Oct4	1
Spl_2i-Esrrb	$Mbd3^{-/-}$	Esrrb	1
Spl_2i-Klf4	$Mbd3^{-/-}$	Klf4	1
Spl_2i-Nanog	$Mbd3^{-/-}$	Nanog	1
Spl_2i-Oct4	$Mbd3^{-/-}$	Oct4	1
Mbd3 2i WT	$Sall4^{Flox/Flox}$	Mbd3	3
Mbd3Null 2i Sall4FLAG	$Mbd3^{-/-}$	Sall4-FLAG	3
S41Null 2i Mbd3	Sall4/1 Double KO	Mbd3	4
S41WT 2i Mi2B	$Sal4^{Flox/Flox}Sall1^{Flox/Flox}$	Chd4	2
Sall4Flag 2i WT	Wild type	Sall4-FLAG	3
Sall4Flag SL WT	Wild type	Sall4-FLAG	3
Sall4Null 2i Mbd3	Sall4 KO	Mbd3	5
S41WT 2i Mbd3	$Sal4^{Flox/Flox}Sall1^{Flox/Flox}$	Mbd3	2

RNAseq datasets produced in this study

Dataset	Cell line	Condition	Replicates
2lox_2i	$Mbd3^{Flox/-}$	2i/LIF	1
2lox_N2B27	$Mbd3^{Flox/-}$	24H N2B27	1
Spl2_2i	$Mbd3^{-/-}$	2i/LIF	1
Spl2_N2B27	$Mbd3^{-/-}$	24H N2B27	1
3KO_2i	$Mbd3^{-/-}$	2i/LIF	1
3KO_N2B27	$Mbd3^{-/-}$	24H N2B27	1
3Flox_2i	$Mbd3^{Flox/-}$	2i/LIF	1
3Flox_N2B27	$Mbd3^{Flox/-}$	24H N2B27	1
S41C2_2i	Sall1 KO	2i/LIF	2
S41C5_2i	Sall1 KO	2i/LIF	2
S41C2_48h_N2B27	Sall1 KO	48H N2B27	2
S41C5_48h_N2B27	Sall1 KO	48H N2B27	2
2A_2i	Sall4 KO	2i/LIF	2
3b2_2i	Sall4 KO	2i/LIF	2
CC4_2i	Sall4 KO	2i/LIF	2

2A_48h_N2B27	Sall4 KO	48H N2B27	2
3B2_48h_N2B27	Sall4 KO	48H N2B27	2
CC4_48h_N2B27	Sall4 KO	48H N2B27	2
S41E8_2i	Sall4/1 DKO	2i/LIF	3
S41F5_2i	Sall4/1 DKO	2i/LIF	3
S41E8_48h_N2B27	Sall4/1 DKO	48H N2B27	2
S41F5_48h_N2B27	Sall4/1 DKO	48H N2B27	2
C7M_2i	$Sall4^{Flox/Flox}$	2i/LIF	2
S4FF1_2i	$Sall4^{Flox/Flox}$	2i/LIF	2
C7M_48h_N2B27	$Sall4^{Flox/Flox}$	48H N2B27	2
S4FF1_48h_N2B27	$Sall4^{Flox/Flox}$	48H N2B27	2
S41FF2_2i	$Sall4^{Flox/Flox}Sall1^{Flox/Flox}$	2i/LIF	3
S41FF2_48h_N2B27	$Sall4^{Flox/Flox}Sall1^{Flox/Flox}$	48H N2B27	2

Gene expression analyses: qPCR

Taqman probes used for single cell analyses (*Ppia, Gapdh, Atp5a1, Esrrb, Pou5f1, Klf2, Zfp42, Klf4, Nanog, Sox1, Sema6a, Foxg1, Map2, Sox4*). Other Taqman probes used: *Foxa2, Cxcr4, T(Brachyury), Sox17, Pax6*. All Taqman probes were from Invitrogen.

Primers for SYBR green:

Ppia 5'-AGCCATGGAGCGTTTTGGGTCC-3' 5'-TGCGAGCAGATGGGGTAGGGA-3'

Map2 5'-GAGAAACAGCAGAGGAGGTA-3' 5'-ATTAGCTGTTTCTTCGGCTG-3'

TuJ1 5'-TGCGRGTGTACAGGTGAATGC-3' 5'-AGGCTGCATAGTCATTTTCCAAG-3'

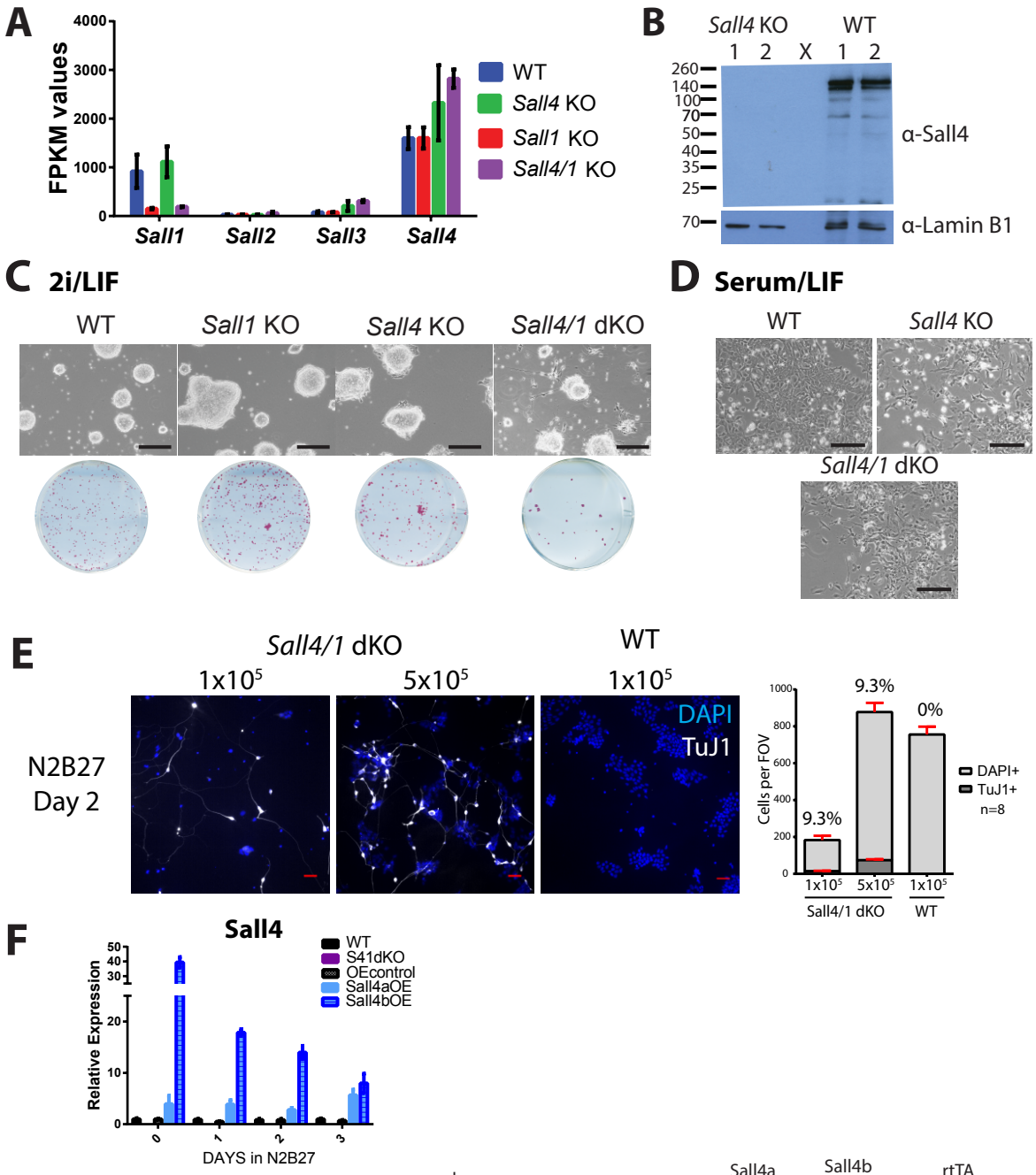
Gfap 5'-TTTCTCCAACCTCCAGATCC-3' 5'-CCCGCATCTCCACAGTCTT-3'

Table S1. Differentially expressed genes: WT vs Sall1 KO, Sall4 KO, Sall41 DKO and Mbd3KO in 2i/LIF and N2B27 conditions

Click here to Download Table S1

Table S2. Closest genes to Sall4 and Mbd3 ChIP-seq peaks in WT cells, and Sall4-bound (and not Mbd3-bound) genes found in Mbd3 KO cells only.

Click here to Download Table S2



Supplemental Figure 1.

(A) FPKM values from RNA-seq analysis showing the expression of the four *Sall* transcripts in WT, *Sall1* null, *Sall4* null and *Sall4/1* double null cells in 2i/LIF. Error bars represent standard deviation between replicates. In 2i/LIF WT cells n=6; *Sall1* null cells n=4; *Sall4* null cells n=6; *Sall4/1* double null n=6; all from ≥2 independent cell lines.

(B) Western blot showing two of the independent *Sall4* null and WT cell lines probed for α-Sall4 and α-LaminB1. The Sall4 antibody used (Abcam ab29112) would recognise a truncated Sall4 protein (which should be ~20kDa) were it present.

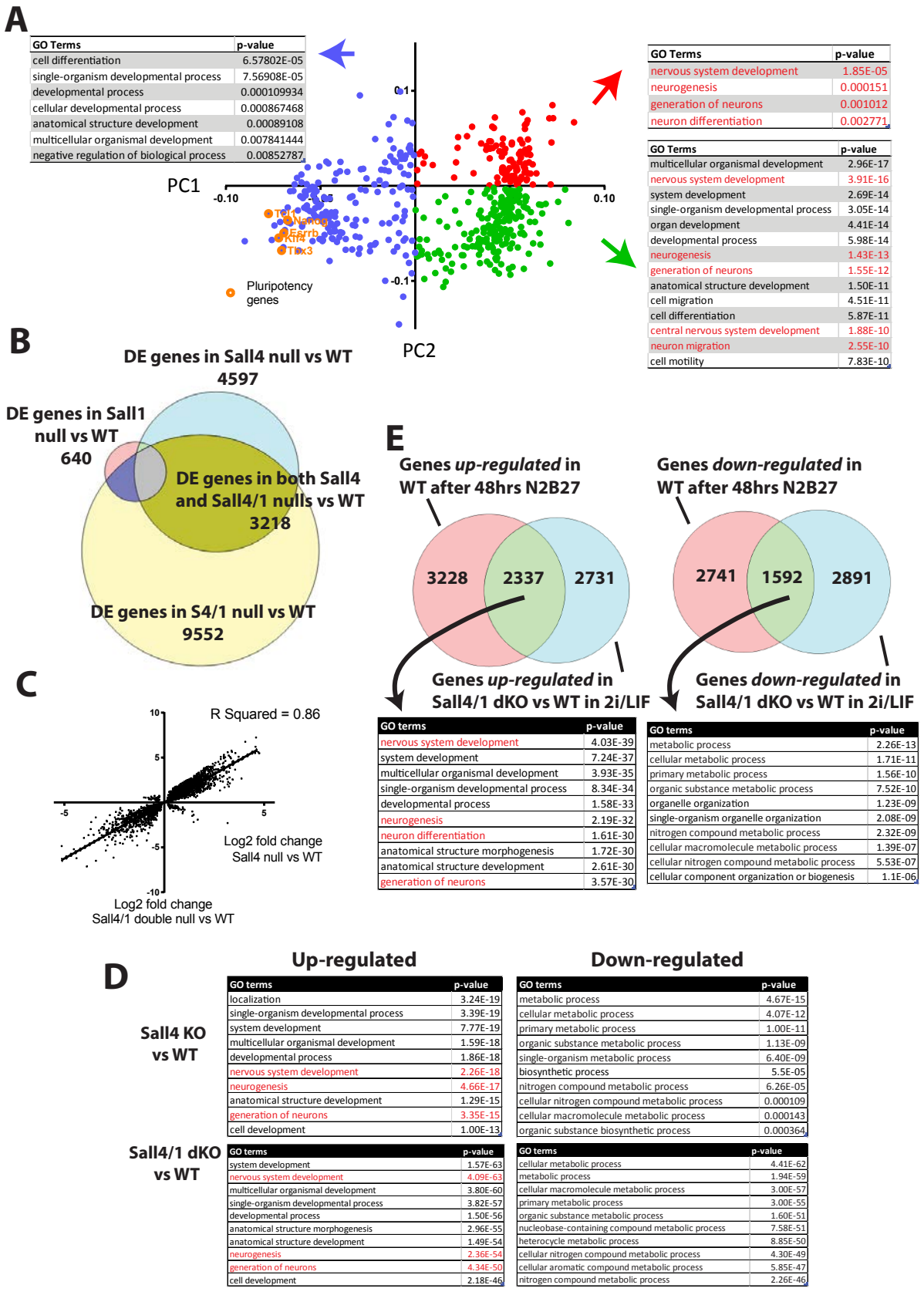
(C) Upper images: Phase contrast images of all cell lines in 2i/LIF conditions: WT, *Sall1* null, *Sall4* null, and *Sall4/1* double null ES cells. Scale bar = 200 μM. Lower images: Alkaline phosphatase staining of each cell line. Cells were plated at 500 cells per 6-well and cultured for 5 days in 2i/LIF, except for the Sall4/1 double null cells that were cultured for 10 days in 2i/LIF before fixing and staining.

(D) Phase contrast images of WT, *Sall4* null, *Sall4/1* double null ES cells after conversion from 2i/LIF to Serum/LIF conditions. Images of WT and *Sall4* null cells were taken after 6 days in Serum/LIF (including one passage) while *Sall4/1* double null cells were taken after 2 days in Serum/LIF conditions (after one passage). Scale bar = 200 μM.

(E) Left images: Representative immunofluorescence images of Sall4/1 double null cells and WT cells stained with α-TuJ1 antibody and DAPI after 2 days in N2B27. The number of cells seeded on Day 0 on laminin per 6-well are shown above the image. The middle and right images are those shown in Fig. 2A. Right graph: Quantification of the images showing the total number of cells (by DAPI staining) per field of view (FOV) and the number of TuJ1 positive cells. Eight images were counted per condition. The percentage shown above each bar shows the percentage of TuJ1+ cells per condition.

(F) Expression of *Sall4* in WT, *Sall4/Sall1* dKO, ES cells overexpressing Sall4a or Sall4b and their control at Day 0, 1, 2 and 3 in N2B27. OE control refers to WT cells expressing the Tet-transactivating factor only. OE control and Sall4OE cell lines were cultured in DOX for the entire timecourse. The Day 0 timepoint was treated the same as Day 1 for all cells except they were culturing in 2i/LIF instead of N2B27. Relative expression refers to the expression relative to housekeeping genes as well as to their respective WT controls.

(G) Western blot showing the overexpression of Sall4a and Sall4b in the Tet-On overexpression cell lines. Hours + DOX indicates the addition of doxycycline for the number of hours shown, all in 2i/LIF conditions. The blot was probed for α-Sall4 and α-Mbd3. The rtTA panels are shown again at a longer exposure to highlight the very low level expression of Sall4b in ES cells. Size markers in KDa are shown at right. The ponceau stained gel image (lower panel) serves as a loading control.



GO Terms	p-value
cell differentiation	6.57802E-05
single-organism developmental process	7.56908E-05
developmental process	0.000109934
cellular developmental process	0.000867468
anatomical structure development	0.00089108
multicellular organismal development	0.007841444
negative regulation of biological process	0.00852787

GO Terms	p-value
nervous system development	1.85E-05
neurogenesis	0.000151
generation of neurons	0.001012
neuron differentiation	0.002771

GO Terms	p-value
multicellular organismal development	2.96E-17
nervous system development	3.91E-16
system development	2.69E-14
single-organism developmental process	3.05E-14
organ development	4.41E-14
developmental process	5.98E-14
neurogenesis	1.43E-13
generation of neurons	1.55E-12
anatomical structure development	1.50E-11
cell migration	4.51E-11
cell differentiation	5.87E-11
central nervous system development	1.88E-10
neuron migration	2.55E-10
cell motility	7.83E-10

GO terms	p-value
nervous system development	4.03E-39
system development	7.24E-37
multicellular organismal development	3.93E-35
single-organism developmental process	8.34E-34
developmental process	1.58E-33
neurogenesis	2.19E-32
neuron differentiation	1.61E-30
anatomical structure morphogenesis	1.72E-30
anatomical structure development	2.61E-30
generation of neurons	3.57E-30

GO terms	p-value
metabolic process	2.26E-13
cellular metabolic process	1.71E-11
primary metabolic process	1.56E-10
organic substance metabolic process	7.52E-10
organelle organization	1.23E-09
single-organism organelle organization	2.08E-09
nitrogen compound metabolic process	2.32E-09
cellular macromolecule metabolic process	1.39E-07
cellular nitrogen compound metabolic process	5.53E-07
cellular component organization or biogenesis	1.1E-06

Up-regulated

GO terms	p-value
localization	3.24E-19
single-organism developmental process	3.39E-19
system development	7.77E-19
multicellular organismal development	1.59E-18
developmental process	1.86E-18
nervous system development	2.26E-18
neurogenesis	4.66E-17
anatomical structure development	1.29E-15
generation of neurons	3.35E-15
cell development	1.00E-13

GO terms	p-value
system development	1.57E-63
nervous system development	4.09E-63
multicellular organismal development	3.80E-60
single-organism developmental process	3.82E-57
developmental process	1.50E-56
anatomical structure morphogenesis	2.96E-55
anatomical structure development	1.49E-54
neurogenesis	2.36E-54
generation of neurons	4.34E-50
cell development	2.18E-46

Down-regulated

GO terms	p-value
metabolic process	4.67E-15
cellular metabolic process	4.07E-12
primary metabolic process	1.00E-11
organic substance metabolic process	1.13E-09
single-organism metabolic process	6.40E-09
biosynthetic process	5.5E-05
nitrogen compound metabolic process	6.26E-05
cellular nitrogen compound metabolic process	0.000109
cellular macromolecule metabolic process	0.000143
organic substance biosynthetic process	0.000364

GO terms	p-value
cellular metabolic process	4.41E-62
metabolic process	1.94E-59
cellular macromolecule metabolic process	3.00E-57
primary metabolic process	3.00E-55
organic substance metabolic process	1.60E-51
nucleobase-containing compound metabolic process	7.58E-51
heterocycle metabolic process	8.85E-50
cellular nitrogen compound metabolic process	4.30E-49
cellular aromatic compound metabolic process	5.85E-47
nitrogen compound metabolic process	2.26E-46

Supplemental Figure 2

(A) The 500 most variable genes which have the biggest impact on the PCA plot (Figure 3A) are plotted with their epigenvector loading values for PC1 and PC2. Gene

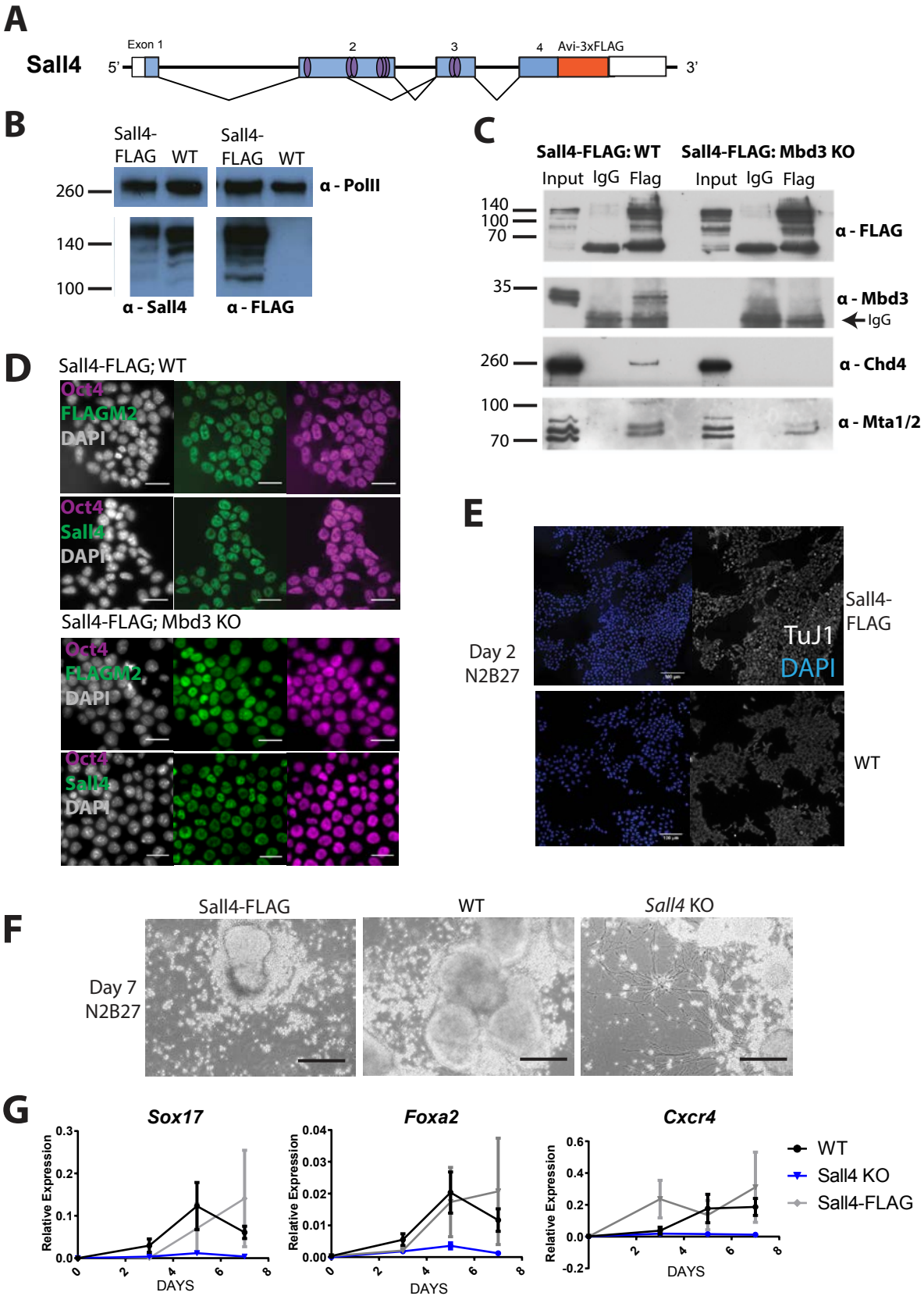
Ontology (GO) analysis was then performed on the red/green and blue sections, of which the top results are shown. A cut off p-value of 0.01 was used for GO terms. Genes associated with pluripotency are labelled in orange. Red GO terms are associated with neural differentiation.

(B) Venn diagram showing the overlap between differentially expressed genes between *Sall1* KO, *Sall4* KO and *Sall4/1* dKO all compared to WT cells. All cells cultured in 2i/LIF conditions.

(C) Log2 fold change is plotted for the differentially expressed genes between *Sall4* KO vs WT, as well as between *Sall4/1* dKO vs WT cells, all in 2i/LIF conditions. Linear regression was performed to generate the R square value.

(D) GO analysis of genes up- and down-regulated in *Sall4* KO vs WT, and *Sall4/1* vs WT cells in 2i/LIF conditions. Red terms are associated with neural differentiation.

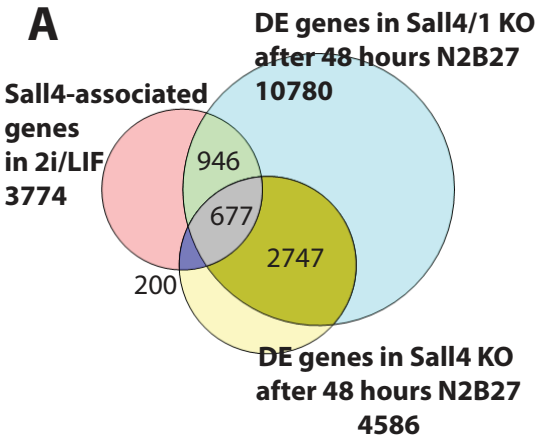
(E) LEFT: Venn diagram showing the overlap between the up-regulated genes that change between WT cells in 2i/LIF and WT cells after 48 hours N2B27 (pink circle) and the genes up-regulated in *Sall4/1* dKO cells vs WT in 2i/LIF (blue circle). RIGHT: Venn diagram showing the overlap between the down-regulated genes that change between WT cells in 2i/LIF and WT cells after 48 hours N2B27 (pink circle) and the genes down-regulated in *Sall4/1* dKO cells vs WT in 2i/LIF (blue circle). The genes overlapping in both comparisons were subjected to GO analysis, and red highlighted terms are associated with neural differentiation.

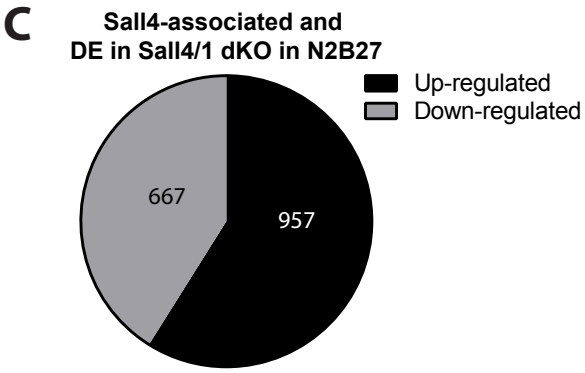


Supplemental Figure 3

(A) Schematic of the Sall4 genomic locus (as in Figure 1A) with the position of the epitope tag (Avi-3xFLAG) shown in red.

(B) Western blot of nuclear extract from Sall4-FLAG and parental (WT) ES cells probed with anti-Sall4 (left), anti-FLAG (right), or anti-RNA Polymerase II (S5P) as a loading control.
(C) FLAG Immunoprecipitation in Sall4-FLAG WT and Sall4-FLAG; *Mbd3* KO ES cell lines. The western blot was probed for α-FLAG, α-Mbd3, α-Chd4 and α-Mta1/2. Size markers are shown at left in KDa.
(D) Immunofluorescence of Sall4-FLAG WT and Sall4-FLAG; *Mbd3* KO ES cell lines stained with α -Sall4, α -Oct4 or α -FLAG antibodies and DAPI in 2i/LIF conditions. Scale bars = 25μm.
(E) Immunofluorescence of Sall4-FLAG WT cells staining for TuJ1 and DAPI after 2 days in N2B27. Scale bar = 100 μm.
(F) Phase contrast images of Sall4-FLAG WT cells after 7 days in N2B27 compared to *Sall4* KO and WT cells. Scale bar = 200 μm.
(G) Expression analysis across the endodermal differentiation time course for the homozygous Sall4-FLAG cells in comparison to WT and *Sall4* KO cells for *Sox17, Foxa2* and *Cxcr4*. Error bars represent SEM between replicates. N=3-9.





B

Up-regulated genes

2i/LIF

GO term	p-value
system development	4.86E-23
multicellular organismal development	1.70E-22
nervous system development	4.76E-22
anatomical structure development	1.92E-21
regulation of multicellular organismal process	3.86E-21
single-organism developmental process	6.08E-20
developmental process	6.36E-20
neurogenesis	6.94E-19
regulation of multicellular organismal development	9.34E-19
generation of neurons	1.75E-18

48 hours N2B27

GO term	p-value
nervous system development	9.39E-29
generation of neurons	3.99E-24
neurogenesis	9.09E-24
movement of cell or subcellular component	1.15E-20
cell development	1.48E-20
neuron differentiation	2.55E-19
multicellular organismal development	2.73E-19
system development	3.51E-19
cell differentiation	1.17E-18
cellular developmental process	3.08E-18

Down-regulated genes

2i/LIF

GO term	p-value
organelle organization	1.03E-08
cellular component organization	1.84E-08
macromolecular complex subunit organization	2.22E-08
cellular component organization or biogenesis	3.86E-08
single-organism organelle organization	2.84E-06
cell cycle	0.003471
chromosome organization	0.004133
chromatin organization	0.004619
primary metabolic process	0.004861
chromatin modification	0.005646

48 hours N2B27

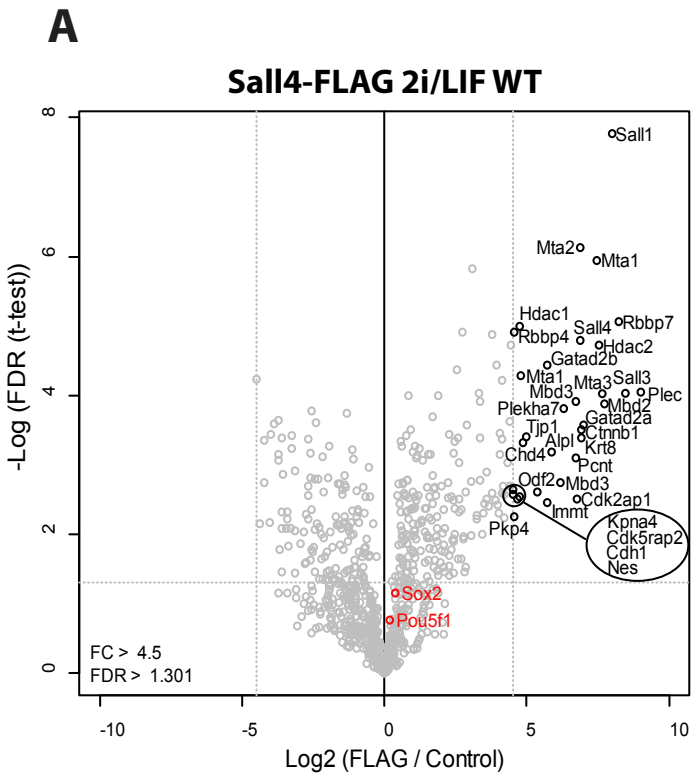
GO term	p-value
macromolecular complex subunit organization	3.41E-05
chromosome organization	0.001377
cellular component organization	0.001712
organelle organization	0.002192
chromatin organization	0.002927
single-organism metabolic process	0.003278
cellular component organization or biogenesis	0.004192
chromatin modification	0.013632
covalent chromatin modification	0.015026
nitrogen compound metabolic process	0.023046

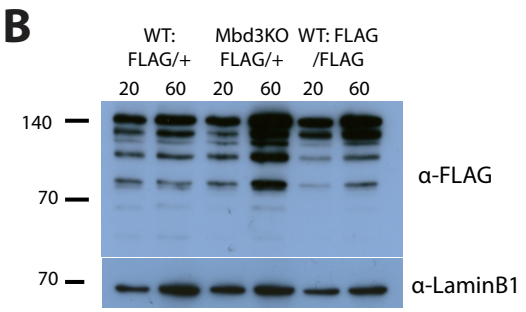
Supplemental Figure 4

(A) Venn diagram showing the overlap between Sall4-associated genes in 2i/LIF (pink circle), differentially expressed genes in *Sall4* KO vs WT after 48 hours N2B27 (yellow circle) and differentially expressed genes in *Sall4/1* dKO compared to WT after 48 hours N2B27 (blue circle).

(B) GO analysis of the genes shown in Figure 4D and Figure S5C. Top results are shown, and the red terms are associated with neural differentiation.

(C) Pie chart showing the Sall4-associated genes (in 2i/LIF) that are differentially expressed in *Sall4/1* dKO cells compared to WT cells, and the proportion that are up-regulated as well as down-regulated after 48 hours in N2B27

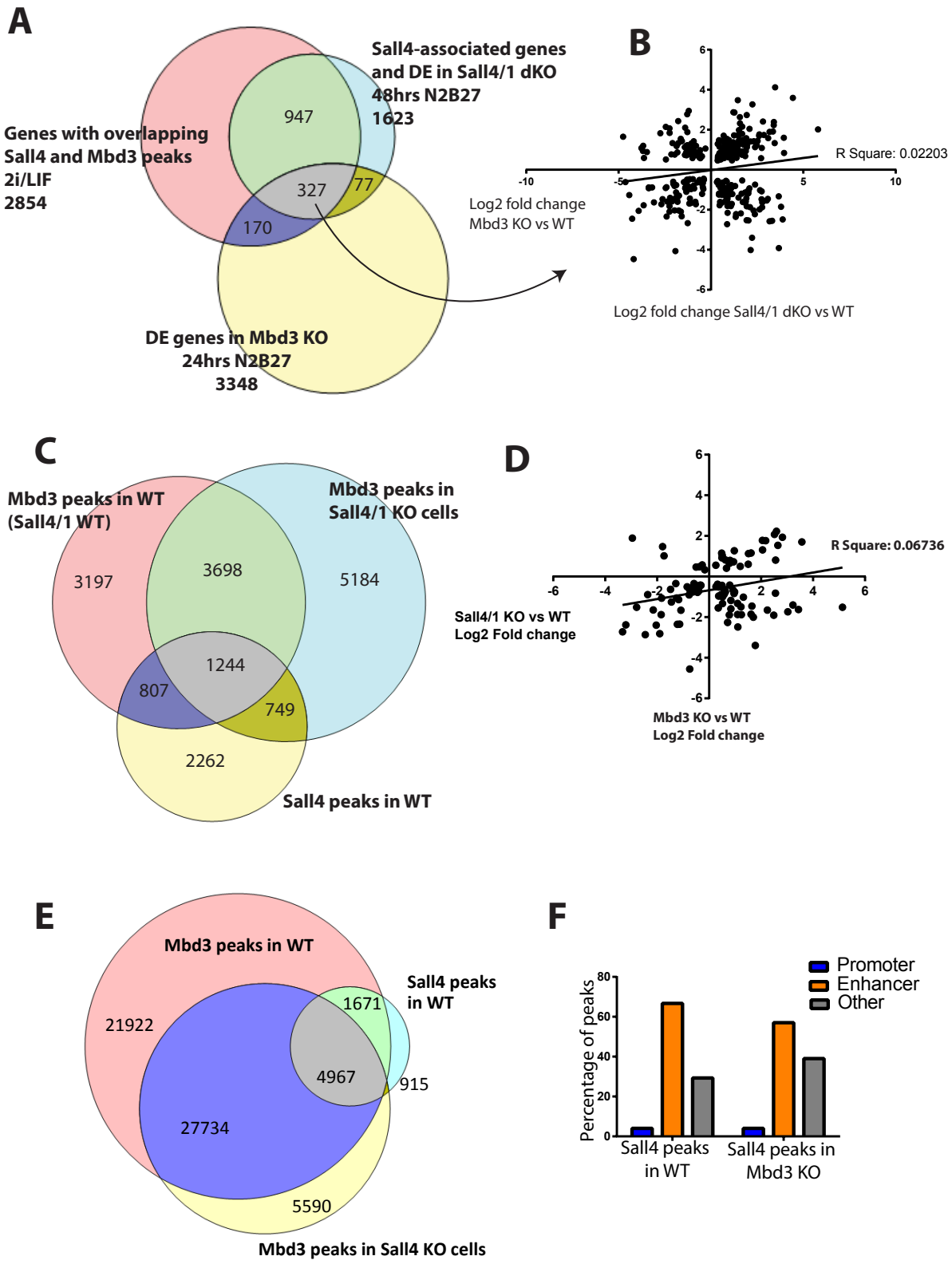




Supplemental Figure 5

(A) Volcano plot showing significant interactors of Sall4 (black circles) in WT ES cells cultured in 2i/LIF conditions. The location of peptides derived from Oct4 (Pou5f1) and Sox2 are shown in red.

(B) Western blot of 20 µg or 60 µg of nuclear extract *from* $Sall4^{FLAG/+}$ $Mbd3^{+/+}$ ES cells (left), $Sall4^{FLAG/+}$ $Mbd3^{-/-}$ ES cells (Middle), and $Sall4^{FLAG/FLAG}$ $Mbd3^{+/+}$ ES cells (right) probed with anti-Sall4 antibody (top) or anti-LaminB1as a loading control.



Supplemental Figure 6

(A) Venn diagram showing the overlap between: pink circle- genes associated with overlapping Sall4 and Mbd3 peaks in 2i/LIF, blue circle- genes associated with a Sall4 peak and differentially expressed in *Sall4/1* KO after 48 hours N2B27 and yellow circle- differentially expressed genes in *Mbd3* KO cells after 24 hours N2B27.

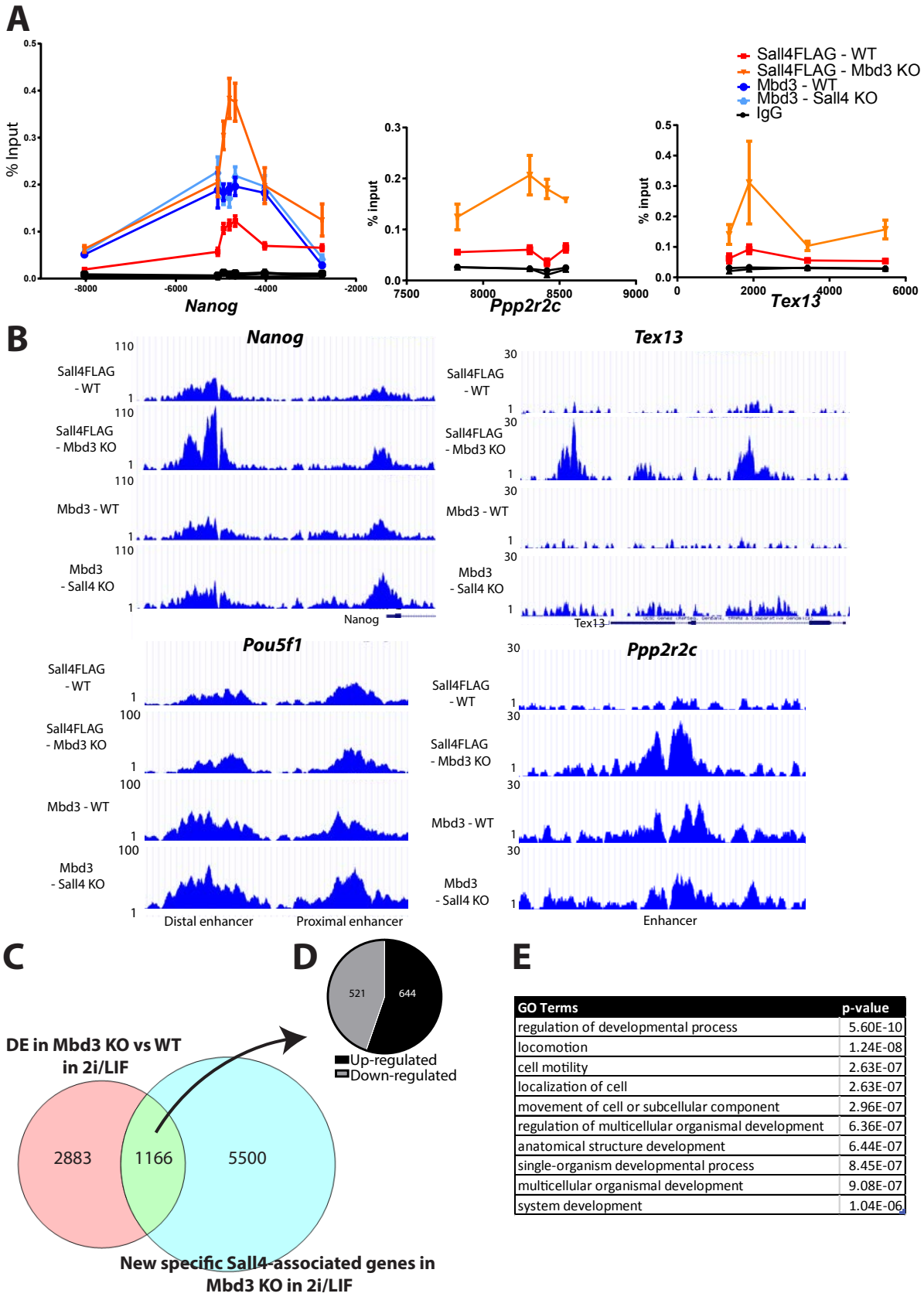
(B) Plot comparing the Log2 fold change between differentially expressed genes in both *Sall4/1* dKO and *Mbd3* KO cells compared to WT cells. The genes are those that have Sall4 and Mbd3 overlapping peaks and that are differentially expressed in Sall4/1 null cells as well as *Mbd3* KO cells compared to WT, all in N2B27 (the 327 genes shown in Figure S6A). A linear regression was performed to generate the R square value.

(C) Venn diagram showing the overlap between: pink circle- Mbd3 peaks in WT cells, blue circle – Mbd3 peaks in *Sall4/1* dKO cells and yellow circle – Sall4 peaks in WT cells. All in 2i/LIF.

(D) Plot comparing the Log2 fold change between differentially expressed genes showing Mbd3 and Sall4 association in wild type cells, and a loss of Mbd3 binding in *Sall4/1* double null ES cells. A linear regression was performed to generate the R square value.

(E) Venn diagram showing the overlap between: pink circle- Mbd3 peaks in WT cells, blue circle – Mbd3 peaks in *Sall4* KO cells and yellow circle – Sall4 peaks in WT cells. All in 2i/LIF. This analysis was using all pooled peaks between replicates instead of the IDR method.

(F) Sall4 peaks in WT cells and in *Mbd3* KO cells broken down into promoter- and enhancer-associated and other peaks. Category definitions can be found in Supplemental Experimental Procedures.



E

GO Terms	p-value
regulation of developmental process	5.60E-10
locomotion	1.24E-08
cell motility	2.63E-07
localization of cell	2.63E-07
movement of cell or subcellular component	2.96E-07
regulation of multicellular organismal development	6.36E-07
anatomical structure development	6.44E-07
single-organism developmental process	8.45E-07
multicellular organismal development	9.08E-07
system development	1.04E-06

Supplemental Figure 7

(A) ChIP-qPCR for indicated proteins in wild type (WT), Mbd3KO or Sall4 KO cell lines across the enhancer elements of *Nanog*, *Tex13* and *Ppp2r2c*. IgG pulldowns were performed as controls. All in 2i/LIF. X-axes indicate base pairs relative to annotated transcriptional start sites.

(B) ChIP-seq traces for Sall4FLAG ChIP in WT and *Mbd3* KO cells, and Mbd3 ChIP in WT and *Sall4* KO cell lines across enhancer elements. Nanog is used as an example of an enhancer that is already bound by Sall4 in WT cells and shows increased enrichment in *Mbd3* KO cells. Tex13 and Ppp2r2c show novel Sall4 binding in the *Mbd3* KO cells, and Pou5f1 is used as an example to show no change between enrichment between WT and *Mbd3* KO cell lines. Mbd3 enrichment does not change in any of these examples between WT and *Sall4* KO ES cells.

(C) Venn diagram showing the overlap between differentially expressed genes in *Mbd3* KO cells vs WT (pink circle) and the genes that have Sall4-associated peaks only in the *Mbd3* KO cells where there is no enrichment for Mbd3 in WT cells, i.e. the genes associated with the 10547 peaks in Figure 6D (blue circle). All in 2i/LIF.

(D) Pie chart showing the proportion of the 1166 genes that are up-regulated and down-regulated in *Mbd3* KO cells vs WT. These genes are those that are associated with a novel Sall4 peak in *Mbd3* KO cells (which do not have significant Mbd3 enrichment in WT cells) that are differentially expressed in *Mbd3* KO cells vs WT. All in 2i/LIF.

(E) Ten most significant GO terms for the 644 upregulated genes indicated in panel (D). Down regulated genes produced no significant ($p \leq 0.01$) GO terms.