

# Enforcement of developmental lineage specificity by transcription factor Oct1

Zuolian Shen<sup>1†</sup>, Jinsuk Kang<sup>1†‡</sup>, Arvind Shakya<sup>1†§</sup>, Marcin Tabaka<sup>2</sup>, Elke A Jarboe<sup>1</sup>, Aviv Regev<sup>2,3</sup>, Dean Tantin<sup>1\*</sup>

<sup>1</sup>Department of Pathology, University of Utah School of Medicine, Salt Lake City, United States; <sup>2</sup>The Broad Institute of MIT and Harvard, Cambridge, United States; <sup>3</sup>Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, United States

**Abstract** Embryonic stem cells co-express Oct4 and Oct1, a related protein with similar DNA-binding specificity. To study the role of Oct1 in ESC pluripotency and transcriptional control, we constructed germline and inducible-conditional Oct1-deficient ESC lines. ESCs lacking Oct1 show normal appearance, self-renewal and growth but manifest defects upon differentiation. They fail to form beating cardiomyocytes, generate neurons poorly, form small, poorly differentiated teratomas, and cannot generate chimeric mice. Upon RA-mediated differentiation, Oct1-deficient cells induce lineage-appropriate developmentally poised genes poorly while lineage-inappropriate genes, including extra-embryonic genes, are aberrantly expressed. In ESCs, Oct1 co-occupies a specific set of targets with Oct4, but does not occupy differentially expressed developmental targets. Instead, Oct1 occupies these targets as cells differentiate and Oct4 declines. These results identify a dynamic interplay between Oct1 and Oct4, in particular during the critical window immediately after loss of pluripotency when cells make the earliest developmental fate decisions.

DOI: [10.7554/eLife.20937.001](https://doi.org/10.7554/eLife.20937.001)

\*For correspondence: dean.tantin@path.utah.edu

†These authors contributed equally to this work

Present address: <sup>‡</sup>C and C Research Laboratories, Suwon, South Korea; <sup>§</sup>Celgene Corporation, San Diego, United States

Competing interest: See page 22

Funding: See page 23

Received: 25 August 2016

Accepted: 23 May 2017

Published: 24 May 2017

Reviewing editor: Clare Blackburn, MRC Centre for Regenerative Medicine, University of Edinburgh, United Kingdom

© Copyright Shen et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

The mammalian blastocyst inner cell mass (ICM) contains undifferentiated, pluripotent cells capable of generating all tissue lineages of the embryo proper. Cultured embryonic stem cells (ESCs) are derived from these cells and have similar capabilities (*Abranches et al., 2009*). The POU transcription factor Oct4/Pou5f1 is an indispensable component of the regulatory circuitry underlying these properties (*Morey et al., 2015*). It is expressed in the ICM and in ESCs where its loss accompanies differentiation (*Nichols et al., 1998*). Oct4 is also widely used to generate induced pluripotent stem cells (iPSCs) from somatic cells (*Takahashi and Yamanaka, 2006*).

Together with other factors, Oct4 sustains pluripotency by activating 'core' targets such as *Pou5f1* (encoding Oct4 itself) and *Nanog* (*Boyer et al., 2005*). It also maintains 'poised' targets, including developmentally critical transcription regulators, in a silent but readily inducible state (*Bernstein et al., 2006; Meissner et al., 2008*). These genes frequently encode developmentally important transcription factors and are marked with a bivalent chromatin signature defined by the simultaneous presence of H3K4me3 and H3K27me3 (*Azuara et al., 2006; Bernstein et al., 2006; Ku et al., 2008; Pan et al., 2007*).

Oct1/Pou2f1 is a widely expressed protein related to Oct4. The two proteins have similar DNA-binding specificity (*Tantin, 2013*). In somatic cells, it regulates stem cell and immune memory phenotypes (*Maddox et al., 2012; Shakya et al., 2015b*) and is associated with cytotoxic stress resistance, glycolytic metabolism and malignant transformation (*Bellance et al., 2012; Shakya et al., 2009; Tantin et al., 2005*). Oct1 amplification and/or overexpression correlates with tumor aggressiveness in esophageal, gastric, prostate, lung, cervical, and colorectal cancer (*Vázquez-*

**eLife digest** Humans and most other animals are composed of hundreds of different types of cell, including nerve cells, muscle cells and blood cells. Despite performing many different roles, these cells all develop from a single fertilized egg, which divides to make a particular group of cells that when studied in the laboratory are called embryonic stem cells (or ESCs for short).

The ability of a cell to become a different cell type is defined as “potency”. ESCs are unique because they can specialize into any type of cell present in the adult organism, and they are therefore called “pluripotent”. However, as the embryo develops, its ESCs gradually lose their potency, and become more and more specialized. The activity of a great number of genes must be regulated during the transition from pluripotent to specialized cells, and some of the mechanisms involved in this transition are still unclear.

ESCs are known to need a gene-regulating protein called Oct4 to remain pluripotent and Shen, Kang, Shakya et al. now show that a similar protein named Oct1 is essential for their transition to becoming more specialized. When the gene for Oct1 was deleted from mouse ESCs, they behaved largely like “normal” ESCs, but could not properly mature into certain cell types such as heart and nerve cells. Molecular analyses revealed that Oct4 and Oct1 compete to regulate the activity of many common genes with opposing outcomes: Oct4 keeps ESCs pluripotent while Oct1 leads them to specialize. The Oct4 protein is abundant in ESCs and prevails over Oct1, but as the cells mature, the levels of Oct4 drop, and Oct1 takes over in the regulation of their common target genes.

Going forward, a better understanding of how ESCs become specialized will help basic research in the laboratory and allow scientists to tackle new questions about how the human body develops and how our organs work. In the longer-term, these findings might also have applications in the field of regenerative medicine, which aims to repair or replace a person’s cells, tissues or organs to improve their health.

DOI: [10.7554/eLife.20937.002](https://doi.org/10.7554/eLife.20937.002)

*Arreguín and Tantin, 2016*). It is also co-expressed with Oct4 in ESCs (*Okamoto et al., 1990; Rosner et al., 1990*). Oct1-deficient mice undergo implantation but show defects following gastrulation, most prominently in extra-embryonic tissues, where trophoblast stem cell development is arrested and expression of the direct Oct1 target *Cdx2* is defective (*Sebastiano et al., 2010*). Tetraploid complementation bypasses this developmental restriction, allowing embryos to survive to E8.5–9.5 where they die from an embryo-intrinsic block. These embryos are runted, developmentally arrested, and lack beating hearts. (*Sebastiano et al., 2010*). A slightly less severe germline allele dies in mid-gestation and manifests runting, anemia, hemorrhaging, and other defects with variable penetrance (*Wang et al., 2004*).

Here, we show that ESCs lacking Oct1 have no discernable defects when maintained in an undifferentiated state, but that silent, normally poised developmental-specific genes fail to induce properly upon differentiation. Additionally, genes specific for alternative developmental lineages are inappropriately expressed. Most prominently, placenta-specific genes not normally expressed in any ESC-derived lineage are induced, indicating that Oct1 restricts extra-embryonic gene expression in differentiating ESCs. Additionally, these cells show phenotypic defects when differentiated into multiple lineages, form smaller and less differentiated teratomas, and fail to generate chimerism when injected into blastocysts. ChIPseq identifies a group of targets co-bound by Oct1 and Oct4 in ESCs associated with non-classical binding sites termed MOREs (More Palindromic Octamer Related Elements, ATGCATATGCAT). These sites are inducibly bound by Oct1 in somatic cells lacking Oct4. The function of Oct1 at these genes is to insulate their expression against repression by oxidative stress, and consistently Oct1-deficient ESCs are hypersensitive to oxidative stress. Oct1 associates with developmentally poised targets upon differentiation and Oct4 loss, explaining the altered gene expression observed with RNAseq. These results establish Oct1 as a key mediator of both developmental-specific gene induction and repression, and identify a dynamic interplay in which Oct1 replaces Oct4 at target genes as ESCs differentiate and early decisions about induction or repression of lineage-specific genes are made.

## Results

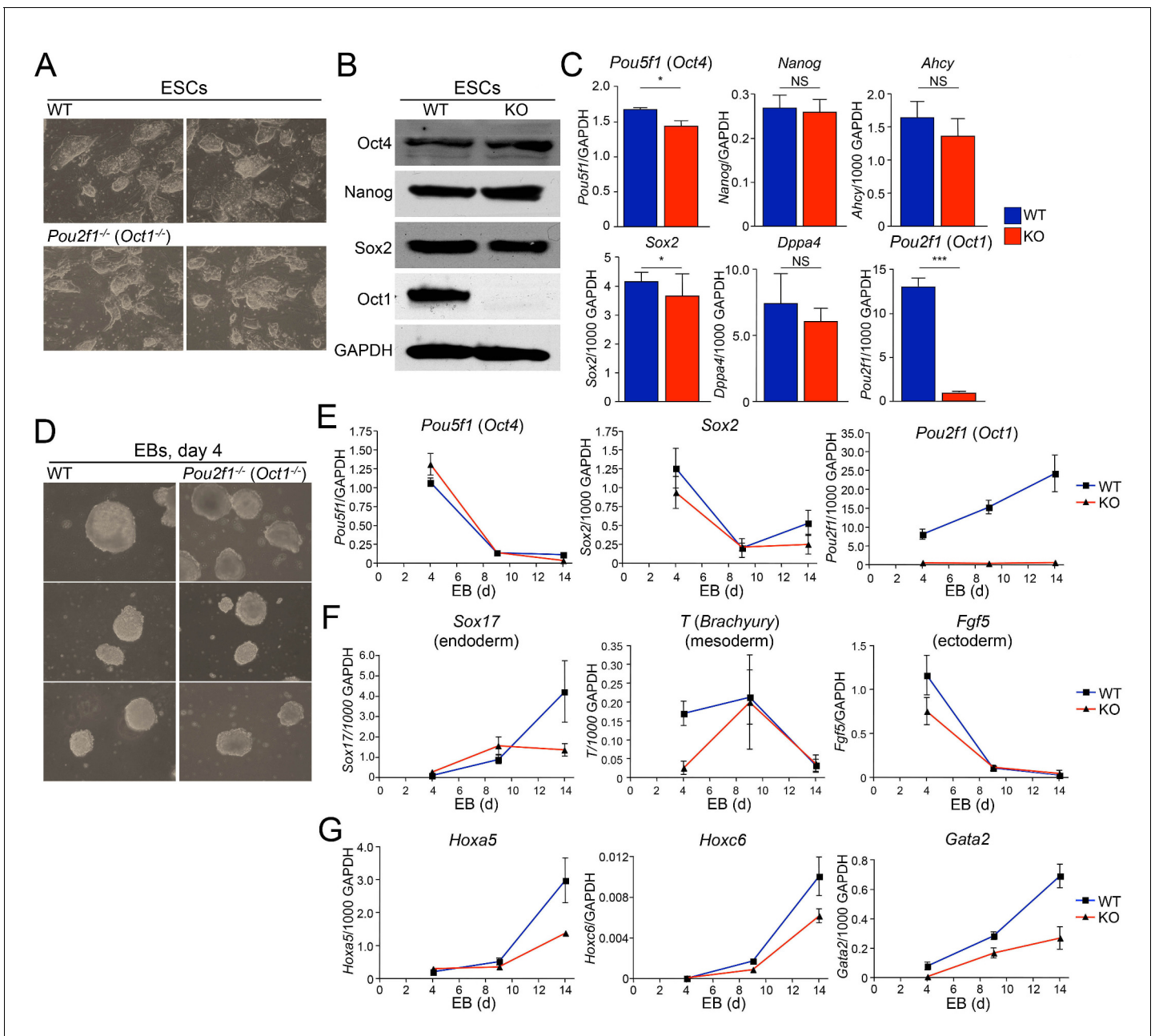
### Oct1 germline-deficient ESCs are phenotypically normal but differentiate abnormally

We derived Oct1-deficient ESC lines by intercrossing *Pou2f1* germline heterozygotes (Wang et al., 2004). Oct1-deficient animals die in utero (Sebastiano et al., 2010; Wang et al., 2004), but survive long enough to derive ESCs. Two Oct1-deficient lines and two littermate WT controls were generated. All had normal karyotypes (not shown). Oct1-deficient ESCs proliferate at normal rates (not shown), are morphologically normal (Figure 1A) and can be propagated for a month in culture with no loss of ESC morphology (not shown). They express normal levels of Oct4, Sox2, and Nanog protein but no Oct1 (Figure 1B). In addition, cells express the pluripotency-associated *Pou5f1* (Oct4), *Sox2*, *Nanog*, and *Dppa4* mRNAs at normal levels (Figure 1C). *Ahcy*, a stress-inducible Oct1 target in which the function of Oct1 is to prevent stress-associated repression (Kang et al., 2009; Shakya et al., 2011), was also unaltered.

To study differentiation, we used early-passage Oct1-deficient and WT control ESCs to form embryoid bodies (EBs). Oct1-deficient ESCs were able to aggregate into EBs at day four with morphology similar to WT (Figure 1D). Similar results were obtained at days 1 and 2 (Figure 1—figure supplement 1A). During EB formation, *Pou5f1* and *Sox2* were down-modulated with similar kinetics in Oct1-deficient and WT cells, while *Pou2f1* (Oct1) remained undetectable (Figure 1E). *Sox17* (endoderm), *Brachyury* (*T*, definitive mesoderm), and *Fgf5* (definitive ectoderm) expression in Oct1-deficient EBs was grossly similar to WT at some (days 4, 9, or 14) timepoints (Figure 1F), consistent with findings that Oct1 is dispensable for gastrulation (Sebastiano et al., 2010; Wang et al., 2004). However, there were consistent defects in expression in the Oct1-deficient condition at day 14 for *Sox17* and day 4 for *T* and *Fgf5*. *Sox17*, *T* and *Fgf5* are known Oct4 targets (Chen et al., 2008). By day 5, Oct1-deficient EBs were somewhat smaller in appearance (Figure 1—figure supplement 1B). We therefore looked for further evidence of defects in induction kinetics in three other known silent but developmentally inducible Oct4 target genes: *Hoxa5*, *Hoxc6*, and *Gata2* (Chen et al., 2008). Each of these genes showed a similar pattern of defective induction in Oct1-deficient EBs relative to WT controls (Figure 1G).

To study gene induction using a more developmentally restricted system, we analyzed expression of known developmentally inducible Oct4 target genes during RA-mediated differentiation of WT and Oct1-deficient ESCs. RA treatment of ESCs ultimately results in a largely neuronal phenotype, but waves of gene expression, differentiation, proliferation, and cell death take place during the course of RA treatment (Walker et al., 2007). Upon differentiation, ESCs  $\pm$ Oct1 lose their clustered, spherical, refractile morphology with similar kinetics (not shown). *Pou5f1* and *Sox2* were also lost with similar kinetics  $\pm$ Oct1, while *Pou2f1* was not detectable in KO cells (Figure 2A). To study developmental gene expression, we tested *Hoxa5*, *Hoxc6*, *Cdx2*, and *Sox17*. These genes encode developmentally important transcription factors and are known Oct4 targets (Chen et al., 2008), but are silent in ESCs. Upon RA-mediated differentiation, lineage-appropriate (ectoderm) genes such as *Hoxa5* and *Hoxc6* (Jiang et al., 2011) normally 'resolve' their bivalent state by losing H3K27me3 and becoming induced, while lineage-inappropriate (e.g. endoderm) genes such as *Cdx2* and *Sox17* normally resolve by losing H3K4me3, gaining DNA methylation, and becoming stably silenced. Induction of *Hoxa5* and *Hoxc6* was robust following RA-mediated differentiation of WT cells, but defective in the Oct1 KO condition (Figure 2B). In contrast, *Cdx2* was ectopically activated upon RA-mediated differentiation of Oct1-deficient ESCs (Figure 2B). Similarly, the definitive endoderm-specific gene *Sox17* is not normally induced upon RA-mediated differentiation, but showed ectopic expression in the absence of Oct1 (Figure 2B). The ectopic *Sox17* expression observed with RA differentiation differed from the expression defects observed in Oct1-deficient EBs, which include endodermal lineages and which showed 100-fold stronger *Sox17* expression (Figure 1F). These results indicate that Oct1-deficient ESCs induce lineage-appropriate developmental genes poorly, while ectopically expressing lineage-inappropriate genes.

In order to determine whether forced Oct1 expression during differentiation was sufficient to correct defects in gene expression, we differentiated Oct1-deficient ESCs using RA and infected the cell during the differentiation timecourse with lentiviral vectors expressing Oct1 and a puromycin resistance cassette, or empty vectors containing the puromycin resistance cassette alone. We confirmed that cells transduced with this vector overexpressed Oct1 by immunoblotting (Figure 2C). Cells



**Figure 1.** Abnormal developmental gene induction in ESCs lacking Oct1. (A) Phase microscopy images of four ESC lines (two Oct1 deficient, two WT littermate controls) derived from *Pou2f1*<sup>+/-</sup> intercrosses. Passage 5 ESCs on feeder fibroblasts are shown. (B) Immunoblot comparing lysates of a WT control line and littermate Oct1-deficient line. GAPDH is shown as a loading control. (C) mRNA expression of six genes in WT control and littermate Oct1-deficient ESC lines. Data were obtained by RT-qPCR using three biological replicates of a single line of each genotype. Error bars denote standard deviations. *p*-values: NS=non-significant, \* < 0.05, \*\* < 0.01, \*\*\* < 0.001. (D) Phase microscopy images of 4-day EBs derived from ESCs ±Oct1. Three representative images of each genotype from wells of a 96-well plate are shown. (E) EBs were collected at 4, 9, and 14 days, and cDNA was prepared and subjected to RT-qPCR. Expression levels were normalized to GAPDH. Pluripotency genes (*Pou5f1*, *Sox2*) and *Pou2f1* were tested. Three biological replicates were performed. Error bars denote ±standard deviation. (F) Additional genes representative of all three germ layers, *Sox17*, *T*, and *Fgf5*, were tested as in E. (G) Three known poised Oct4 target genes, *Hoxa5*, *Hoxc6*, and *Gata2*, were tested as in E.

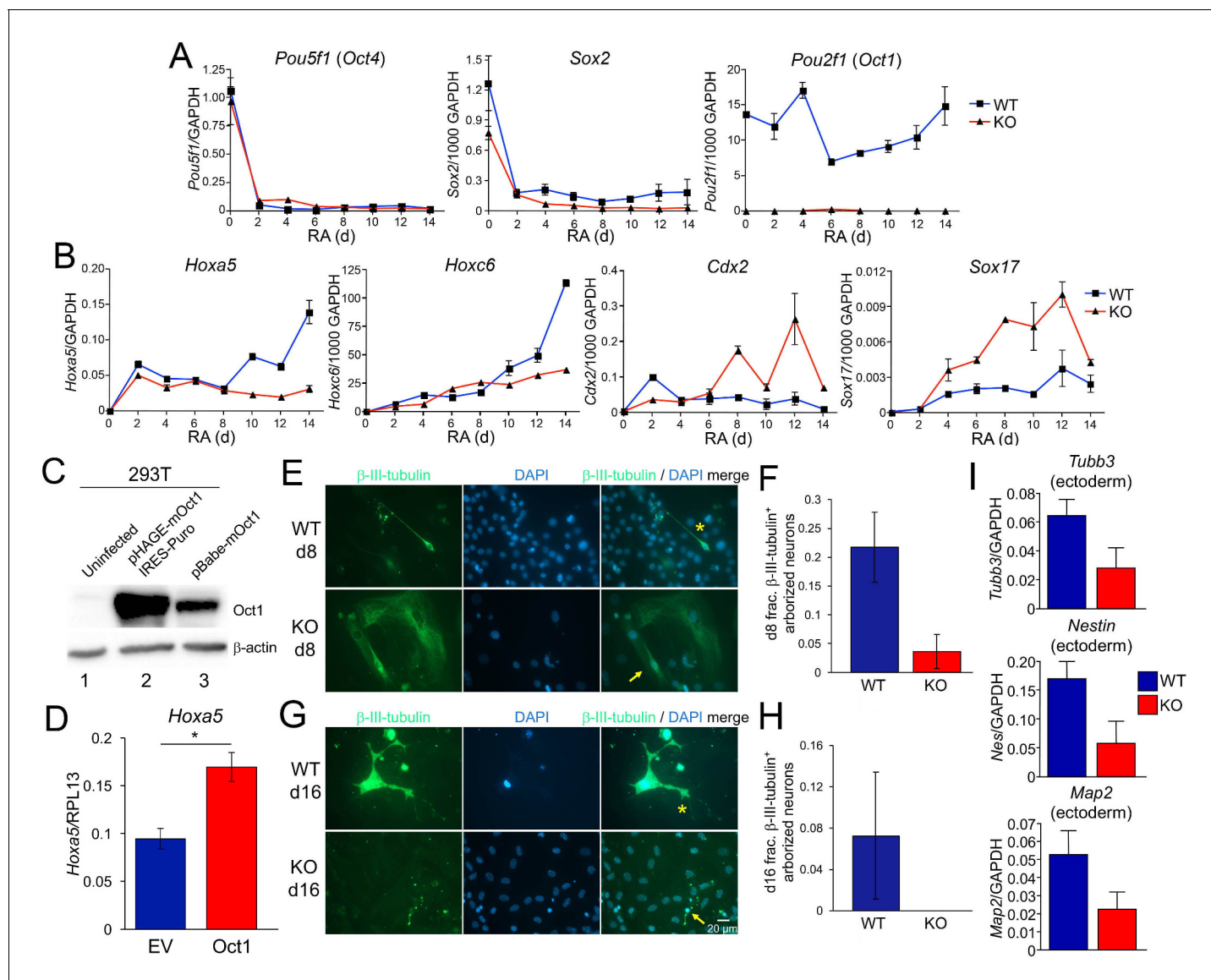
DOI: 10.7554/eLife.20937.003

The following figure supplement is available for figure 1:

**Figure supplement 1.** Abnormal morphology in differentiating Oct1-deficient cells manifests by day 5 of EB formation.

DOI: 10.7554/eLife.20937.004





**Figure 2.** Effect of Oct1 loss on RA-mediated differentiation and neurogenesis. (A) Quantitative RT-PCR results are shown for *Pou5f1* (Oct4), *Sox2*, and *Pou2f1* (Oct1) mRNA relative to a GAPDH standard. Average of three biological replicates  $\pm$  standard deviation is shown. (B) Similar analysis performed for the Oct4 targets *Hoxa5*, *Hoxc6*, *Cdx2*, and *Sox17*. (C) 293 T cells were transiently transfected with a lentiviral vector (pHAGE) expressing mouse Oct1. Lysates were prepared 48 hr later and immunoblotted for Oct1. Un-transfected cells are shown as a negative control (lane 1). Oct1 is not visible in these cells because of the lightness of the exposure. The same cells transduced with a retroviral vector encoding Oct1 (lane 3) are shown as a positive control.  $\beta$ -actin is shown as a loading control. (D) Oct1-deficient ESCs were differentiated using RA for 14 days. 4 days into the timecourse, cells were infected with lentiviruses expressing Oct1 and a puromycin resistance cassette, or an empty vector (EV) control. Cells were selected with puromycin for the remainder of the timecourse. cDNAs from the endpoint cultured were used to study expression of *Hoxa5* relative to an RPL13 ribosomal protein internal standard. Cells were prepared in triplicate for each condition. Error bars denote  $\pm$  standard deviation. \*denotes  $p < 0.05$ . (E) Immunofluorescence images of WT and Oct1-deficient ESCs differentiated into neurons. Cells were cultured as EBs for 8 days, followed by culture for a further 8 days in neuralizing media (see Materials and methods).  $\beta$ -tubulin III and DAPI staining are shown. (F) Quantification of 300–400 cells from three individual differentiation experiments. Error bars denote  $\pm$  standard deviation. (G) Similar to (E) except cells were cultured for eight additional d in neuralizing media. (H) Similar to (F) except using cells cultured for eight additional d. (I) 16 d-differentiated neuron cultures of similar genotypes were pooled and subjected to RT-qPCR using primers specific for *Tubb3*, *Nestin*, and *Map2*. Expression was assessed relative to GAPDH. Averages of three biological replicates are shown. Error bars denote  $\pm$  standard deviation.

DOI: 10.7554/eLife.20937.005

The following figure supplement is available for figure 2:

**Figure supplement 1.**  $\beta$ -tubulin III staining of neuralizing WT and Oct1-deficient EBs.

DOI: 10.7554/eLife.20937.006

were infected over 2 consecutive days and selected with puromycin throughout the remainder of the 14-day differentiation timecourse. At timepoints after day 6, infection and selection with empty vector skewed the expression of genes such as *Hoxa5*, suggesting that infection and selection were skewing the populations of cells in the culture. Cells infected at 4 and 5 days, however, did not show major differences (not shown), suggesting that the composition of cells in culture was not being significantly altered. We therefore infected differentiating Oct1-deficient cells consecutively on days 4 and 5, prepared cDNAs at day 14 and examined gene expression. By RT-qPCR, Oct1 was undetectable in cells transduced and selected with empty vector but robustly expressed by cells transduced with Oct1 (not shown). Expression of the developmentally-inducible *Hoxa5* gene was significantly augmented ( $p=0.026$ ) by ectopic Oct1 expression (**Figure 2D**). These results indicate that restoration of Oct1 expression at these times and conditions can correct at least some of the gene expression defects associated with Oct1 deficiency.

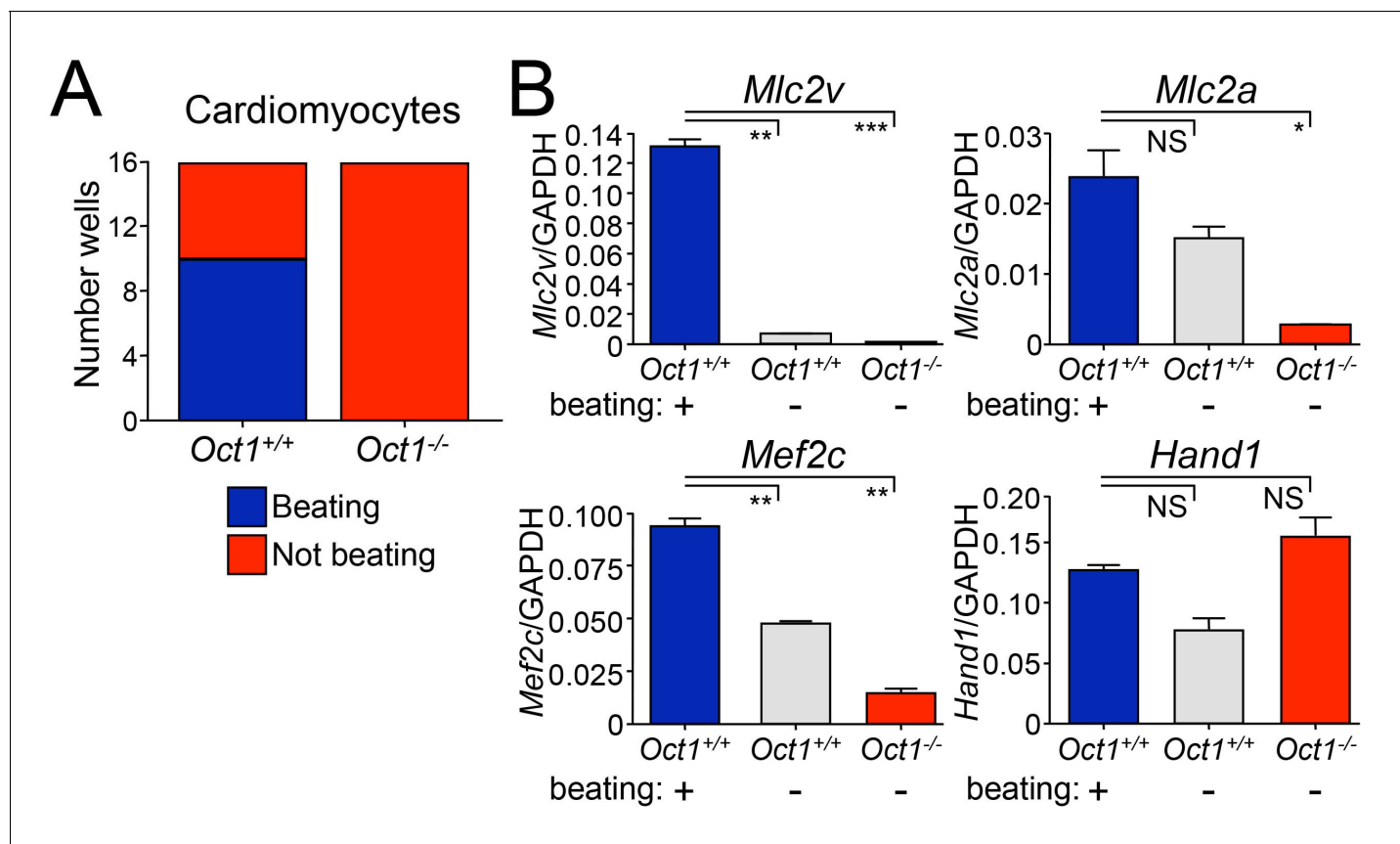
RA-mediated differentiation yields neuronal precursor cells but not neurons. We used a differentiation system involving EB generation and culture in insulin, transferrin and selenium (see Materials and methods) to generate arborized neurons that express the marker  $\beta$ -tubulin III (*Tubb3*) and the neuroectoderm genes Nestin (*Nes*) and *Map2*. Staining of 2-day-old EBs for  $\beta$ -tubulin III prior to laminin/poly-L-lysine dish attachment – early in the differentiation protocol – revealed fewer  $\beta$ -tubulin III-positive cells in the Oct1 deficient condition (**Figure 2—figure supplement 1**). Upon complete differentiation (8 d EBs, 8 days in neuralizing monolayer culture), WT ESCs formed neurons robustly (**Figure 2E**, asterisk) while few  $\beta$ -tubulin III-expressing neurons were formed from Oct1-deficient ESCs. Oct1-deficient cells that did induce  $\beta$ -tubulin III tended to do so at lower levels, and the few cells that did express  $\beta$ -tubulin III robustly were nevertheless abnormal (**Figure 2E–F**, arrow). To test if Oct1 loss induced a kinetic delay that could be overcome by longer culture, cells were incubated for 8 or 16 additional days (16 or 24 days in neuralizing medium, 24 or 32 days total differentiation). In neither case were neurons formed (**Figure 2G–H** and data not shown). To study gene expression, individual wells of common genotypes differentiated for 16 days were pooled and subjected to RT-qPCR for *Tubb3*, *Nes* and *Map2*. Each of these genes showed defective expression in the absence of Oct1 (**Figure 2I**).

To test an unrelated developmental system, we performed cardiomyocyte differentiation by culturing EBs in hanging drops followed by culture with gelatin (see Materials and methods). Oct1-deficient ESCs failed to form beating cardiomyocytes, unlike WT (**Figure 3A** and **Videos 1–12**). RNA was collected from pooled beating and non-beating WT colonies, and Oct1-deficient colonies, and used to analyze *Mef2c* and *Hand1*, regulators of cardiomyocyte differentiation, and the terminal differentiation markers *Mlc2v* and *Mlc2a*. We observed *Mlc2v* expression defects in the Oct1-deficient condition equivalent to non-beating cardiomyocyte colonies from WT ESCs. *Mlc2a* and *Mef2c* expression was even weaker than non-beating WT cardiomyocyte colonies (**Figure 3B**). In contrast, *Hand1* showed no expression defects (**Figure 3B**), indicating that Oct1 deficiency does not globally down-regulate genes associated with cardiomyocyte differentiation. Cumulatively, the results indicate that although Oct1-deficient ESCs appear normal in the absence of differentiation cues, they do not induce poised developmentally inducible genes and fail to repress lineage-inappropriate genes such as *Cdx2* and *Sox17*, resulting in multiple cellular defects following differentiation.

## Oct1 conditional-deficient ESCs display abnormal gene expression upon differentiation

Although the ESC lines described above were derived from littermate animals and had normal karyotypes, it was possible that the developmental phenotypes and altered gene expression patterns resulted from differences unrelated to Oct1 status. Furthermore, the observed gene expression defects could result from compensatory changes due to development in an Oct1-deficient environment. Finally, the allele used to generate these lines is a severe hypomorph rather than a complete null (**Wang et al., 2004**). To circumvent these issues, we generated tamoxifen-inducible, Oct1 conditional-deficient ESCs.

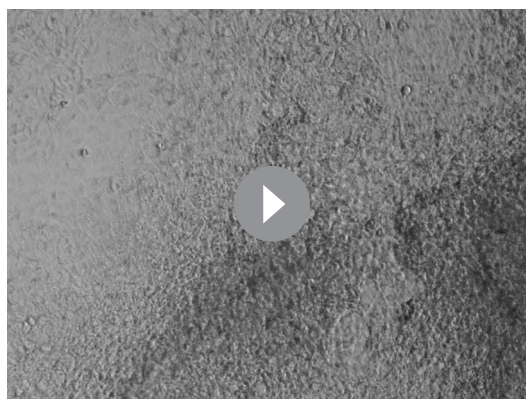
We previously described *Pou2f1* conditional (floxed) mice (**Shakya et al., 2015b**). We generated inducible-conditional Oct1 ESCs by crossing the floxed allele onto *Rosa26-Cre-ERT2* and *Rosa26-lox-stop-lox-YFP* (see Materials and methods). Pregnant animals were used to isolate *Pou2f1<sup>fl/fl</sup>; Rosa26-Cre-ERT2; Rosa26-lox-stop-lox-YFP* ESC lines in which Oct1 could be acutely deleted and YFP induced by 4-hydroxytamoxifen (4-OHT) administration. Treatment of parent ESCs with 4-OHT



**Figure 3.** Defective cardiomyocyte differentiation in ESCs lacking Oct1. (A) Cardiomyocytes were generated from individual EBs using 24-well dishes with gelatin. Functionality ( $\pm$ beating) was assessed for each well (16 per genotype) and plotted. (B) The wells assessed in (A) were pooled according to genotype and function (beating WT, non-beating WT and non-beating Oct1 deficient), cDNA was prepared and used for RT-qPCR using primers for *Mlc2v*, *Mlc2a*, *Mef2c* and *Hand1*. Averaged results from three replicates are shown. Error bars denote standard deviation. p-Values: NS=non-significant, \* < 0.05, \*\* < 0.01, \*\*\* < 0.001.

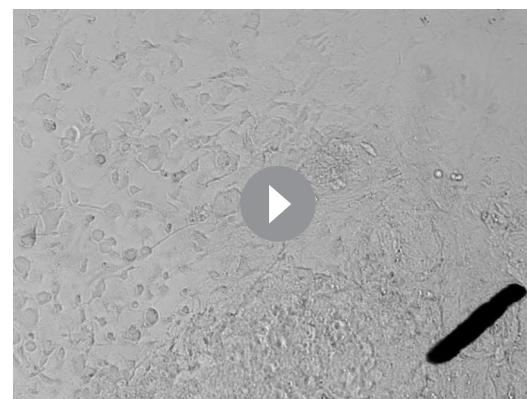
DOI: 10.7554/eLife.20937.007

resulted in variegated YFP<sup>+</sup> colonies (Figure 4A, step one at top). Colonies with good morphology were picked (red arrow), trypsinized and expanded into derived ESC lines (Figure 4A, step two at



**Video 1.** Example WT ESC line cardiomyocyte differentiation 1.

DOI: 10.7554/eLife.20937.008



**Video 2.** Example WT ESC line cardiomyocyte differentiation 2.

DOI: 10.7554/eLife.20937.009



**Video 3.** Example WT ESC line cardiomyocyte differentiation 3.

DOI: [10.7554/eLife.20937.010](https://doi.org/10.7554/eLife.20937.010)

glucose oxidation (Shakya et al., 2009). These changes contribute to failure of fibroblasts to undergo oncogenic transformation, despite the fact that they grow at normal rates and can be immortalized by serial passage. To test if similar changes occur in ESCs lacking Oct1, we analyzed the metabolic profile of these cells. Few differences were noted, with only phosphoethanolamine (also known as phosphorylethanolamine,  $p=0.047$ ) and inositol ( $p=0.037$ ) showing significant changes (Figure 4—figure supplement 1). The lack of difference may be due to redundant functions of co-expressed Oct4 and Oct6, or co-selection for metabolic stability when selecting and propagating ESCs.

Derived *Pou2f1<sup>Δ/Δ</sup>* ESC lines, and parent cell line controls, were subjected to RA-mediated differentiation. Similar to results using Oct1 germline-deficient ESCs, the derived *Pou2f1<sup>Δ/Δ</sup>* ESCs lost Oct4 and Sox2 expression with kinetics identical to the parent line (Figure 4E). Microscopic imaging of the differentiating cells revealed that they were morphologically similar until approximately d 12, at which point *Pou2f1<sup>Δ/Δ</sup>* cells showed an increase in columnar/epithelial appearance (Figure 4—figure supplement 2). Also as before, the induction of silent, developmentally poised genes was defective: *Hoxa5* and *Hoxc6* both showed reduced expression in timecourse assays (Figure 4F). The cells also showed ectopic *Cdx2* expression upon RA treatment (Figure 4F). As with germline-deficient ESCs, *Pou2f1<sup>Δ/Δ</sup>* ESCs did not generate true neurons efficiently (Figure 4G).

To determine the effect of conditional Oct1 loss during differentiation, parent cells were treated with 4-OHT following 8 d EB formation and 4 d in insulin, transferrin and selenium. After an additional 4 days, cells were fixed and stained with antibodies against  $\beta$ -tubulin III to score neurogenesis and YFP to score deletion. 40–50% of the treated cells induced YFP. Nearly all cells that induced  $\beta$ -tubulin III and/or generated neuron morphology lacked YFP expression (Figure 4I and Figure 4J). A few cells (2/ ~ 700) were both YFP- and  $\beta$ -tubulin III-positive (not shown), though it is possible that these cells are *Pou2f1* heterozygous as 4-OHT treatment can result in recombination of only one allele (Figure 4B).

### Oct1-deficient ESCs form smaller, less differentiated teratomas and fail to generate chimeric mice

Parent and *Pou2f1<sup>Δ/Δ</sup>* ESCs were injected subcutaneously into contralateral flanks of NCr Nude immunocompromised animals to generate teratomas. ESCs lacking Oct1 consistently generated smaller tumors (Figure 5A–C). Immunoblotting confirmed that recovered tumors maintained their original Oct1 status (Figure 5D). Histological analysis confirmed that parent cells generated mature teratomas that included, e.g., glial tissue, and glandular epithelial and squamous elements (Figure 5E). In contrast, Oct1 deficient ESCs generated areas of focally immature cells, consistent with reduced differentiation. Occasionally tumors were comprised virtually entirely of primitive malignant cells resembling a germ cell tumor (Figure 5E, lower right).

bottom) that genotyped as *Pou2f1<sup>Δ/Δ</sup>* or *Pou2f1<sup>fl/Δ</sup>* (Figure 4B). The designation  $\Delta$  will be used to differentiate this allele from the germline deficient allele used in Figures 1–3. As with Oct1 germline-deficient ESCs, derived *Pou2f1<sup>Δ/Δ</sup>* ESC lines displayed normal colony morphology (Figure 4C), proliferated normally (not shown) yet expressed no Oct1 (Figure 4D). The derived cells showed normal karyotype profiles and could be propagated for >1 month without loss of an undifferentiated phenotype (not shown). Similar to germline Oct1 deficient ESCs, derived *Pou2f1<sup>Δ/Δ</sup>* cells also expressed Oct4, Sox2 and Nanog at normal levels (Figure 4D).

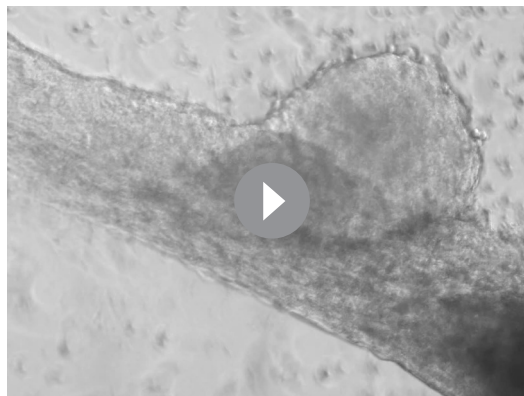
In differentiated cells, Oct1 promotes glycolysis and dampens mitochondrial function. Oct1 deficiency dramatically increases mitochondrial amino acid oxidation and oxygen consumption while decreasing glycolysis and to a lesser extent



A standard measure of pluripotency is the ability to contribute efficiently to adult cells and tissues (De Los Angeles et al., 2015). We injected parent and *Pou2f1<sup>Δ/Δ</sup>* ESCs into albino C57BL/6 blastocysts, resulting in high contribution in the case of the parent line (Figure 5F, left), but no contribution in the case of the derived lines (right). The average percent chimerism from two separate sets of injections (33 animals from parent cell line injections, 36 combined from two different *Pou2f1<sup>Δ/Δ</sup>* lines) confirmed the lack of contribution (Figure 5G). 18/33 animals injected with parent ESCs showed some detectable chimerism (55%), while 1/36 animals injected with conditional knockout ESCs showed transient trace chimerism in the eye (0.03%). The cells were imaged immediately prior to blastocyst injection to confirm an undifferentiated phenotype (Figure 5—figure supplement 1).

### Defects in lineage-specific gene expression in differentiated Oct1-deficient ESCs

To identify gene expression changes stemming from loss of Oct1, we performed RNAseq with undifferentiated and 14 d RA-differentiated parent and *Pou2f1<sup>Δ/Δ</sup>* ESCs. Three independent replicates were performed for each of the four conditions. Between 18.1 and 24.9 million sequence reads were generated for each sample, 73% to 82% of which aligned uniquely to the mouse *Mm10* reference genome. 99.6% of the reads within coding regions aligned to the correct strand. Variance between replicates was similar regardless of genotype or differentiation state (not shown). Unsupervised hierarchical clustering indicated that 0 and 14 days samples separated clearly from each other regardless of genotype, while within each timepoint the KO and parent WT samples clustered together (Figure 6—figure supplement 1A). These results indicated that the effect of RA treatment and differentiation on gene expression was far stronger than the effect of Oct1 deletion. Plotting gene expression levels in the parent vs *Pou2f1<sup>Δ/Δ</sup>* cells (Figure 6A) showed relatively few gene expression changes in the undifferentiated condition (>2.5 fold,  $p < 0.01$ , 253 total genes). These genes never changed by >7 fold (Supplementary file 1). In contrast, 1123 genes change expression in differentiated *Pou2f1<sup>Δ/Δ</sup>* cells, some of which varied by >200 fold. Plotting gene expression fold change vs.  $p$ -value (Figure 6B) recapitulated these findings. Comparing genes differentially expressed at the two timepoints revealed little overlap (23 genes, Figure 6C). Analysis of the genomic alignments revealed that expression of control genes such as *Tbp* was unaltered, while pluripotency genes such as *Nanog* were silenced equivalently (Figure 6D). Other pluripotency genes such as *Pou5f1*, *Klf4*, *Dnmt3l*, and *Dppa4* behaved similarly to *Nanog* (not shown). One gene showing increased expression in the undifferentiated state was *Pou2f3* (Oct11, Figure 6—figure supplement 1B). *Pou2f3* shows low but detectable mRNA expression in WT ESCs, and is repressed upon differentiation to nearly undetectable levels. It is slightly elevated in Oct1-deficient ESCs but decreases to an even greater extent upon differentiation. RT-qPCR confirmed these changes in the context of overall low expression (Figure 6—figure supplement 1C). It is therefore unlikely that this protein provides a compensatory function upon differentiation.



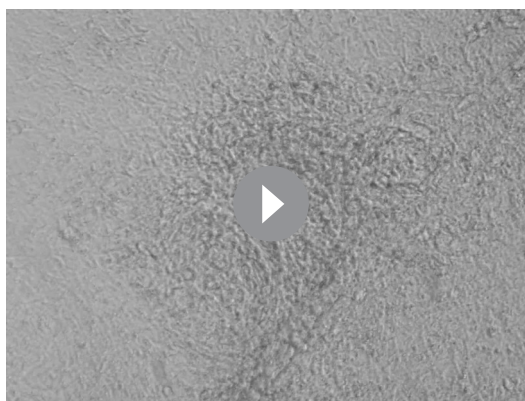
**Video 4.** Example WT ESC line cardiomyocyte differentiation 4.

DOI: 10.7554/eLife.20937.011



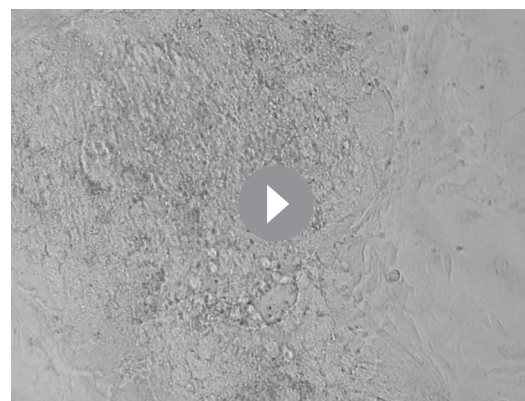
**Video 5.** Example WT ESC line cardiomyocyte differentiation 5.

DOI: 10.7554/eLife.20937.012



**Video 6.** Example WT ESC line cardiomyocyte differentiation 6.

DOI: [10.7554/eLife.20937.013](https://doi.org/10.7554/eLife.20937.013)



**Video 7.** Example WT ESC line cardiomyocyte differentiation 1.

DOI: [10.7554/eLife.20937.014](https://doi.org/10.7554/eLife.20937.014)

Approximately 300 genes were induced poorly in RA treated *Pou2f1<sup>Δ/Δ</sup>* cells relative to normal controls. Many of these encode regulators of neuronal specification and differentiation. Examples include *Hoxa5*, *Hoxb9* and *Nppa* (**Figure 6E**). Other examples include *Foxg1*, *Pcdh17*, *Ptgfr*, *Akr1c18*, *Cts7*, *Duox2*, *Hoxc5*, *Hoxc6*, *Hoxc8*, *Hoxc10*, and *Dcx* (**Figure 6—figure supplement 2A**). In addition, *Ahcy*, a stress-responsive Oct1 target (**Kang et al., 2009; Shakya et al., 2011**) showed weakened expression in the absence of Oct1 specifically in the differentiated condition (**Figure 6E**). An even larger cohort of ~800 genes was aberrantly expressed upon differentiation of *Pou2f1<sup>Δ/Δ</sup>* cells. These genes are strongly associated with alternative developmental fates (**Figure 6F** and **Figure 6—figure supplement 2B**). Examples include *Sox17*, *Cdx2* and *Gata4* (endoderm), *Fgb* (endoderm/liver), *Gata2* (mesoderm/endothelial), *Pparg* and *Irx3* (mesoderm/mesenchymal), *Muc13* (epithelial/hematopoietic) and *Tnfrsf9* (which encodes CD137/hematopoietic). The difference in *Sox17* and *Gata2* expression between EBs (defective) and RA-differentiated cells (elevated) likely arises from the additional developmental fates specified in EBs.

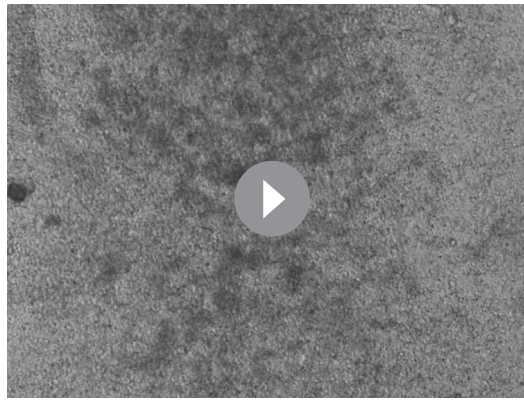
Unexpectedly, differentiating *Pou2f1<sup>Δ/Δ</sup>* cells also resulted in inappropriate expression of genes associated with trophoblast and placental development, the specification of which is normally restricted to trophoblast cells rather than the inner cell mass (from which ESCs are derived). Examples include *Cdx2* (which is also expressed in endoderm), *Prl8a6*, *Hand1* (which is also expressed in cardiomyocytes), *Pappa2*, *Prl3b1*, and *Psg27* (**Figure 6F** and **Figure 6—figure supplement 2B**). Some of these genes are also expressed in other lineages while others are highly specific. In aggregate, they indicate improper activation of an extra-embryonic program. Using RT-qPCR we confirmed unperturbed expression of *Tbp*, defective expression of *Ahcy*, and elevated expression of *Prl8a6* in differentiating Oct1-deficient ESCs (**Figure 6G**). These results indicate that Oct1 deficiency results in defective lineage specification upon differentiation.

We used ChIPseq to identify common and unique Oct1 and Oct4 target genes in ESCs. We also performed H3K4me3 ChIPseq as a control. The ChIPseq data were of high quality based on measures of signal/noise ratio (see Materials and methods). After filtering, 27.3 (Oct1), and 23.7 (Oct4) million alignable reads were generated, corresponding to 692 (Oct1), and 8673 (Oct4) peaks. Allocating the peaks to nearest genes revealed 209 unique Oct1 target genes, 356 common targets, and 5563 unique Oct4 targets (**Figure 7A**). The smaller size of the Oct1 target



**Video 8.** Example WT ESC line cardiomyocyte differentiation 2.

DOI: [10.7554/eLife.20937.015](https://doi.org/10.7554/eLife.20937.015)



**Video 9.** Example WT ESC line cardiomyocyte differentiation 3.

DOI: [10.7554/eLife.20937.016](https://doi.org/10.7554/eLife.20937.016)



**Video 10.** Example WT ESC line cardiomyocyte differentiation 4.

DOI: [10.7554/eLife.20937.017](https://doi.org/10.7554/eLife.20937.017)

pool relative to Oct4 may be attributable to >10 fold lower Oct1 levels in ESCs as observed by RT-qPCR (**Figure 1C**, **Figure 1E**, **Figure 2A**) and RNAseq (not shown). Oct1 may also require a more open chromatin context, and/or the presence of specific co-bound factors, to access DNA. For example *Ii2* and *Ifng* are known Oct1 targets in differentiated T cells (**Shakya et al., 2015b, 2011**) but were not identified as targets in this analysis.

Motif analysis of unique and co-bound peaks revealed significant differences in recognized DNA elements. Regions associated exclusively with Oct4 were significantly enriched for Oct-Sox compound elements that likely also associate with Sox2 in ESCs (**Figure 7B**). In contrast, target regions preferentially associated with Oct1 were enriched for the simple octamer element ATTTGCAT (shown by the software as an Oct4 motif in **Figure 7B**). Interestingly, co-occupied peaks strongly associate with a motif termed a MORE that is known to bind two Oct protein molecules (**Reményi et al., 2001; Tomilin et al., 2000**). In differentiated cells lacking Oct4, oxidative stress induces homodimeric Oct1 binding to MORE-containing genes such as *Polr2a*, *Ahcy*, *Ell*, and *Rras2*. Oxidative stress-induced binding occurs via phosphorylation of a conserved serine residue in the DNA-binding domain (**Kang et al., 2009**). These genes were constitutively co-bound by Oct1 and Oct4 in ESCs (**Figure 7—figure supplement 1**). Additional examples of genes associated with Oct4 alone (*Pou5f1*), or Oct1 alone (*Taf12*) are shown in **Figure 7C**. This panel also shows another example of a MORE containing gene (*Polr2a*, two tandem MOREs binding four molecules) that also associates with both proteins but shows an Oct1 bias, as well as an example (*Pax6*) that is bound by both proteins but in two different locations. Using ChIP-qPCR we validated two genes, *Polr2a* (Oct1-enriched) and *Pou5f1* (Oct4-enriched, **Figure 7D**). The complete set of identified targets is shown in **Supplementary file 2**.

Intersecting the ChIPseq and RNAseq data revealed little overlap. Only 34 Oct1-bound or Oct1/Oct4 co-bound targets showed differential expression following RA-mediated differentiation (**Figure 7A**). Examples include *Pank4*, *Cdh5* and *Med16*. 193 Oct1-bound and 325 Oct1/Oct4-co-bound genes did not show expression differences at d 14. Examples include *Tbx3*, *Tcf4*, and *Txb6*, which also showed no differences throughout the differentiation timecourse (**Figure 7E**). Instead, 1066 genes with altered expression in differentiated Oct1-deficient cells showed Oct4 but not Oct1 enrichment. These findings indicate that (1) identified Oct1 targets were not differentially expressed upon



**Video 11.** Example WT ESC line cardiomyocyte differentiation 5.

DOI: [10.7554/eLife.20937.018](https://doi.org/10.7554/eLife.20937.018)





**Video 12.** Example WT ESC line cardiomyocyte differentiation 6.

DOI: [10.7554/eLife.20937.019](https://doi.org/10.7554/eLife.20937.019)

mRNA specifically in Oct1-deficient ESCs (**Figure 8B**), exactly as observed in fibroblasts (**Kang et al., 2009; Shakya et al., 2011**). As expected, these cells were hypersensitive to H<sub>2</sub>O<sub>2</sub> (**Figure 8C**). These results suggest that as in other cell types, Oct1 functions in ESCs to buffer these genes from oxidative stress-associated inhibition.

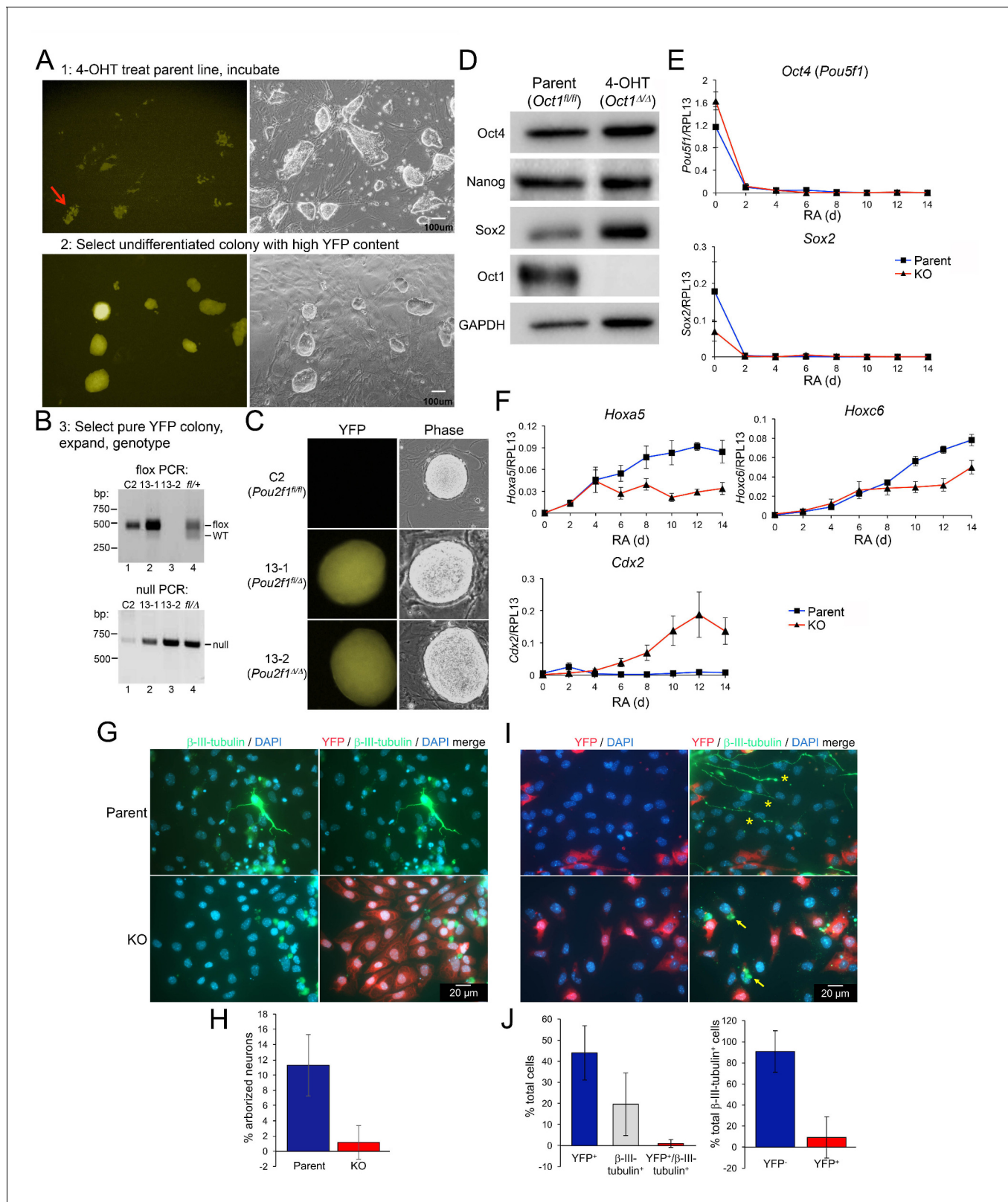
To test the hypothesis that Oct1 occupies Oct4 targets as cells differentiate and Oct4 is lost, we performed ChIP-qPCR timecourses using differentiating ESCs and antibodies against Oct1 and Oct4. Material was collected from 0, 2, 4, 6, 8, 10, 12 and 14 d of differentiation with RA. We chose a gene, *Hoxc5*, that contains a conserved perfect octamer sequence (**Figure 8D**), but is not an Oct1 target based on ChIPseq (**Figure 8E**). *Hoxc5* also shows poor induction in upon RA-mediated differentiation of *Pou2f1*<sup>Δ/Δ</sup> cells (**Figure 6—figure supplement 2A**). Oct1 ChIP-qPCR revealed no binding in ESCs, as expected based on the ChIPseq (**Figure 8F**), however robust binding was transiently observed at 6 d. By 14 d of differentiation Oct1 binding was again undetectable. We also examined a target region between the linked *Myf5* and *Myf6* (*Mrf4*) loci on chromosome 10 that contains several near-perfect octamer sites (not shown), and is strongly bound by Oct4 but not Oct1 (**Figure 8E**). Oct1 inducibly occupied this region even more rapidly (2 d) as Oct4 binding was lost, and in this case Oct1 binding was maintained, at varying levels, during ESC differentiation (**Figure 8F**). Finally, we studied two additional genes, *Pou5f1* and *Rest* (*Nrsf*), which also contain conserved perfect octamer sequences in their regulatory regions and also show exclusive Oct4 binding. These genes are both expressed in ESCs and silenced as differentiation proceeds. Here early Oct1 binding was identified, which was maintained at low levels during the differentiation timecourse in the case of *Pou5f1* but transient in the case of *Rest* (**Figure 8F**). These results indicate a highly dynamic interplay between Oct1 and Oct4 in differentiating cells.

## Discussion

Our results indicate that Oct1-deficient ESCs are unperturbed in terms of morphology, growth, metabolism, and gene expression. EBs formed from these cells are microscopically normal at early timepoints and express genes associated with all three germ layers. However, Oct1-deficient ESCs show phenotypic and molecular defects upon differentiation. These cells fail to form neurons and cardiomyocytes, generate smaller and less differentiated teratomas, and fail to contribute to adult mouse tissues. Prior work has shown that partial knockdown of Oct1 also inhibits neuron formation in the context of knockout of the related protein Oct2 (**Theodorou et al., 2009**).

Three molecular defects manifest upon differentiation of Oct1-deficient ESCs. First, loss of Oct1 results in a failure to fully induce genes associated with a given developmental lineage. Second, Oct1 is necessary for the repression of alternative embryonic developmental lineages. As a result, upon differentiation gene expression programs are marked not only by poor induction of lineage-appropriate gene expression, but also by ectopic expression of genes specific to alternative lineages





**Figure 4.** Gene expression defects upon differentiation of Oct1 inducible-conditional ESCs. (A) YFP-epifluorescence and phase microscopy images of inducible-conditional ESCs. Top: parent *Pou2f1<sup>fl/fl</sup>* cells were treated with 500 nM 4-OHT for 24 hr. A colony with good morphology and variegated YFP expression was picked, trypsinized, replated and expanded. Bottom: derived *Pou2f1<sup>Δ/Δ</sup>* ESCs. (B) PCR genotyping of parent (C2) and derived (13-1, 13-2) lines. Feeder fibroblasts were depleted by two serial 1 hr platings on plastic. The residual WT band in lane two is due to feeder contamination. The Figure 4 continued on next page

Figure 4 continued

rightmost lane (lane 4) shows control tail DNA from a *Pou2f1<sup>fl/+</sup>* (top) or *Pou2f1<sup>fl/Δ</sup>* animal. (C) Epifluorescence and phase microscopy images of single colonies. Images were taken at the same magnification. (D) Immunoblots comparing lysates of vehicle-treated parent ESCs and derived KO cells. GAPDH is shown as a loading control. (E) Feeder-depleted ESCs were treated continuously with RA on gelatin-coated plates in the absence of LIF for 14 days. Media was changed every other day. cDNA was prepared every other day and used in RT-qPCR with primers against *Pou5f1* and *Sox2*. Averages of three biological replicates  $\pm$  standard deviation are shown. Methodologically, the experiments were performed identically to **Figure 2A–B**. (F) Additional RT-qPCR using primers against the Oct4 target genes *Hoxa5*, *Hoxc6* and *Cdx2*. (G) *Pou2f1<sup>Δ/Δ</sup>* ESCs were differentiated into neurons as in **Figure 2C–G** by forming EBs for 8 days followed by 8 day in medium containing insulin, transferrin and selenium. Cells were fixed and used for immunofluorescence using DAPI and antibodies against YFP and  $\beta$ -tubulin III. (H) Immunofluorescence images were quantified based on arborized morphology. Approximately 700 total cells were analyzed. (I) Similar to (G) except parent *Pou2f1<sup>fl/fl</sup>*, ESCs were used and after 4 days in culture cells were treated with 500 nM 4-OHT for 24 hr to delete Oct1 and induce YFP. Two representative images are shown. (J) Immunofluorescence images were quantified based on YFP and  $\beta$ -tubulin III positivity. Percent total cells showing single or double staining, or percent  $\beta$ -tubulin III<sup>+</sup> cells with and without YFP are shown. Approximately 700 total cells were analyzed.

DOI: [10.7554/eLife.20937.020](https://doi.org/10.7554/eLife.20937.020)

The following figure supplements are available for figure 4:

**Figure supplement 1.** Steady-state metabolite levels in parental WT and 4-OHT-treated *Pou2f1<sup>Δ/Δ</sup>* ESCs as determined by GC-MS.

DOI: [10.7554/eLife.20937.021](https://doi.org/10.7554/eLife.20937.021)

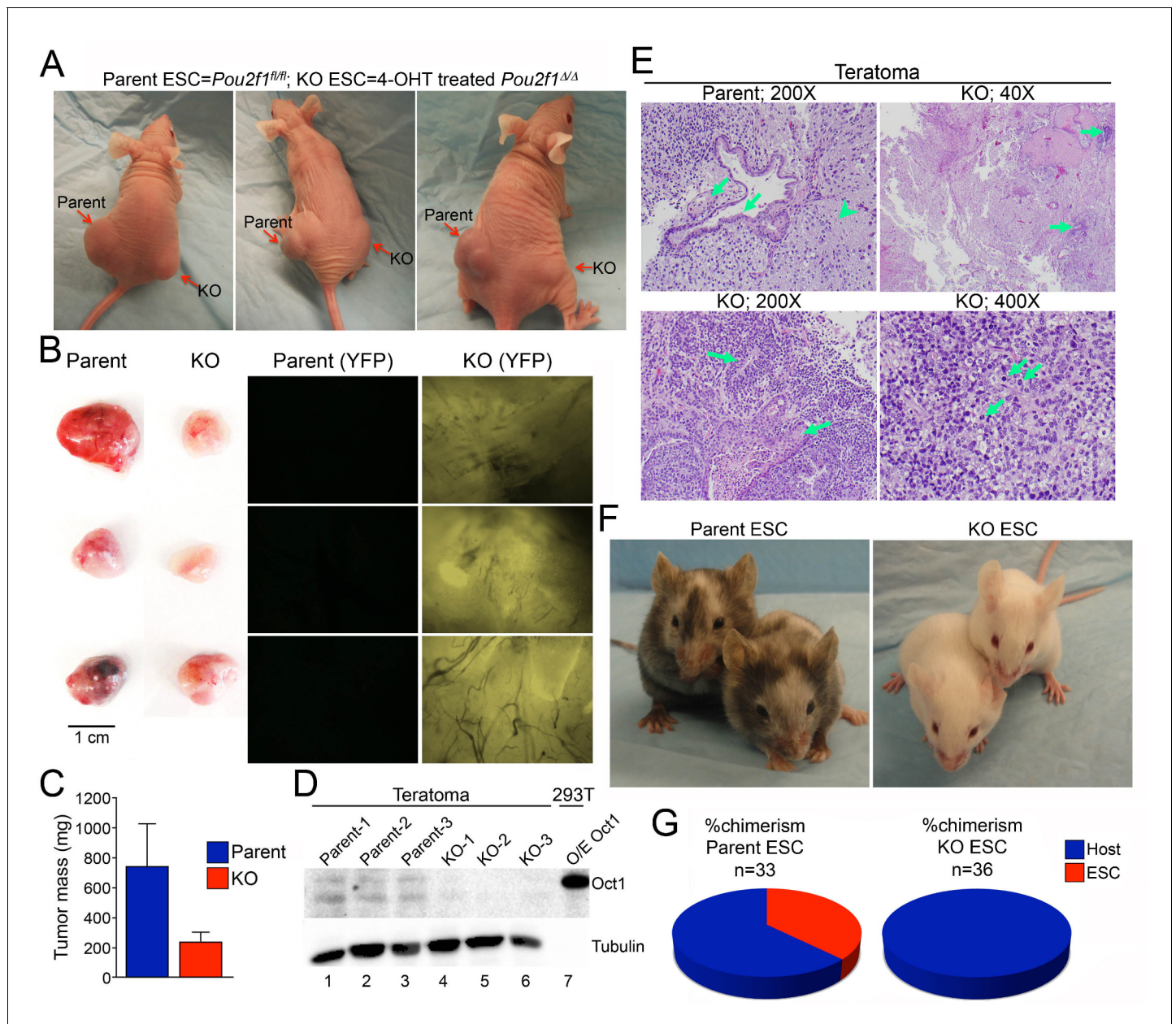
**Figure supplement 2.** Images of parental WT and 4-OHT-treated KO ESCs differentiating in the presence of RA.

DOI: [10.7554/eLife.20937.022](https://doi.org/10.7554/eLife.20937.022)

(**Figure 8G**). The third defect is mis-expression of genes associated with extra-embryonic lineages. These genes are normally under tight repression in ESCs and their differentiated progeny. For example, *Prl8a6* and *Prl3b1* are mis-expressed in RA-differentiated ESCs. Other genes such as *Cdx2* and *Hand1* are examples of genes expressed both in extra-embryonic and alternative embryonic lineages. For example, the *Cdx2* promoter contains a perfect consensus octamer element and is a known Oct1 target in somatic cells (*Jin and Li, 2001; Wang et al., 2009*). In the early embryo, *Cdx2* promotes trophoblast fate and is under tight repression by Oct4 (*Yeap et al., 2009; Yuan et al., 2009*). Later in development, *Cdx2* is induced in the endoderm-derived developing gastrointestinal tract (*Guo et al., 2004; Lu et al., 2008*) and during primitive hematopoiesis (*Wang et al., 2008*), but is not widely expressed in ectoderm (*Suh and Traber, 1996*). In RA-differentiated cells, *Cdx2* thus represents both a lineage-inappropriate gene and an extra-embryonic lineage. *Cdx2* is mis-expressed following RA-mediated differentiation of both germline and inducible-conditional Oct1-deficient ESCs. Interestingly Oct1 may execute the opposite function in extra-embryonic tissue, as germline Oct1-deficient mice show defects in extra-embryonic tissues including poor expression of *Cdx2* (*Sebastiano et al., 2010*).

ChIPseq experiments reveal that Oct1 and Oct4 regulate common and distinct targets in ESCs. These differences in bound targets lead to functional consequences, as the two proteins recruit different cofactors such as Jmjd1a in the case of Oct1 and Jmjd1c in the case of Oct4 (*Shakya et al., 2015a, 2011*). Oct4 occupies a large group of >5000 genes, including developmentally poised genes such as *Hoxc5* and *Myf5*, and core pluripotency genes such as *Pou5f1* and *Nanog*. Oct1 does not occupy these genes in ESCs, consistent with the ability of Oct1-deficient ESCs to maintain pluripotency. Instead Oct1 co-occupies a cohort of 325 genes with Oct4 that are highly enriched for a motif known as a MORE (*Reményi et al., 2001; Tomilin et al., 2000*). Oct proteins are known to homo- and hetero-dimerize (*Kang et al., 2009; Tantin et al., 2008; Tomilin et al., 2000; Verrijzer et al., 1992*). The configuration of Oct proteins can determine cofactor association and hence regulatory output (*Reményi et al., 2001; Tomilin et al., 2000*). Many of these constitutively co-bound genes were previously shown to become occupied by Oct1 upon oxidative stress exposure in differentiated cells lacking Oct4 (*Kang et al., 2009*). The function of Oct1 at these genes is to insulate them against inhibition by oxidative stress. Fibroblasts lacking Oct1 show inappropriate repression of MORE-containing genes following H<sub>2</sub>O<sub>2</sub> exposure (*Kang et al., 2009; Shakya et al., 2011*). We demonstrate the identical phenotype using Oct1-deficient ESCs. Oct1 also exclusively associates with a small number (~200) of other genes including *Taf12*, which contains another binding site variant known as a TMFORE (*Kang et al., 2009*).

Notably, in undifferentiated cells Oct1 does not associate with developmental targets that become deregulated upon differentiation of Oct1-deficient ESCs. Oct4 is present at higher levels in



**Figure 5.** Smaller, less differentiated teratomas and lack of contribution to adult mouse tissues in *Pou2f1<sup>Δ/Δ</sup>* ESCs. (A)  $1 \times 10^6$  ESCs were injected into flanks (left flank: control *Pou2f1<sup>fl/fl</sup>* parent cells, right flank: derived *Pou2f1<sup>Δ/Δ</sup>* ESCs) of NCr Nude mice. Images are shown at 4 weeks. (B) Images of dissected teratomas. Left side: white light. Scale in cm shown on the right. Right side: YFP fluorescence. (C) The mass from ten tumors was averaged and plotted. Error bars denote standard deviation. (D) Immunoblots are shown for Oct1 in lysates prepared from part of the individual teratomas. Lane 7: extracts from 293 T cells transiently over-expressing recombinant Oct1.  $\alpha$ -Tubulin is shown as a loading control. The lack of  $\alpha$ -Tubulin in lane seven arises from the fact that less protein was loaded due to high levels of recombinant Oct1. (E) H and E images of parent and KO teratomas. Top left: normal teratoma morphology comprised of mature elements, e.g glial tissue (arrowhead), mature squamous (left arrow) and ciliated glandular epithelial (right arrow). Top right: teratoma comprised predominantly of mature elements, but with immature elements (approximately 5% of the tumor, arrows). Bottom left: teratoma with both mature and immature elements represented. Mature squamous epithelium (right arrow) is abundant, while immature neuroepithelium (left arrow) is noted focally. Bottom right: tumor comprised almost entirely of a primitive malignant neoplasm that does not recapitulate any recognizable line of differentiation. Arranged in sheets and irregular nests, these cells exhibit marked cytologic atypia, with nuclear pleomorphism and coarsely-clumped chromatin. Nucleoli are variably prominent. Mitotic activity is brisk (arrows). (F) Parent ESCs (left side) or derived KO cells (right side) were injected into albino C57BL/6 blastocysts and implanted in pseudo-pregnant animals. Representative images are shown. (G) Average contribution is shown for the two cell types. ESCs contribution was assessed subjectively based on dark coat and eye color. 33 animals were tested in the case of the parent line and 36 animals were assessed in the case of the derived *Pou2f1<sup>Δ/Δ</sup>* line.

DOI: 10.7554/eLife.20937.023

Figure 5 continued on next page



Figure 5 continued

The following figure supplement is available for figure 5:

**Figure supplement 1.** Pluripotent phenotype of ESCs immediately prior to blastocyst injection.

DOI: 10.7554/eLife.20937.024

ESCs compared to Oct1, suggesting that mass action may contribute to the lack of Oct1 binding. This model predicts that Oct1 would occupy these genes as Oct4 is lost during differentiation. We tested four regions bound by Oct4 but not Oct1 in ESCs, *Hoxc5*, *Myf5/Myf6*, *Rest* and *Pou5f1*, predicting that Oct1 binding will manifest as cells differentiate and Oct4 is lost. In all cases, Oct1 binding was observed at one or more points during the differentiation timecourse. We propose that Oct1 transiently replaces Oct4 at many such Oct4 target genes upon differentiation, where it promotes lineage-appropriate target gene expression, and represses expression of lineage-inappropriate targets. The binding events occur during a brief but important window during which critical decisions about suppression or potentiation of lineage-specific developmental Oct4 target genes are made. Binding also occurs before many of the affected target genes are induced, suggesting that Oct1 is not the principal driver of expression of these genes, but instead establishes a chromatin context in which these genes remain poised for expression, or become permanently repressed. Of the genes tested in ChIP-qPCR RA differentiation timecourses, the lineage-appropriate *Hoxc5* gene shows poor induction in upon differentiation of Oct1-deficient cells (**Figure 6—figure supplement 2A**), *Myf5* and *Myf6* are mesoderm-specific and lineage-inappropriate, *Rest* is both pluripotency-associated and lineage-inappropriate, and *Pou5f1* is more restricted to ESCs. These latter genes showed no evidence of ectopic expression. This observation can be reconciled with our model by positing that redundant mechanisms, perhaps mediated by other Oct proteins such as Oct6/Pou3f1, enforce their repression in differentiating ESCs.

The bipotential function of Oct1 is consistent with previous findings in fibroblasts and T cells (**Shakya et al., 2011**). Oct1 functions are mediated in part through association with the inhibitory chromatin remodeling complex NuRD (**Shakya et al., 2011**), or with Jmjd1a/KDM3A, a histone H3K9me2 lysine demethylase (**Shakya et al., 2015b, 2011**). H3K9me2 also controls developmental gene induction (**Wen et al., 2009; Zyllicz et al., 2015**) and reprogramming to pluripotency (**Chen et al., 2013; Sridharan et al., 2013**).

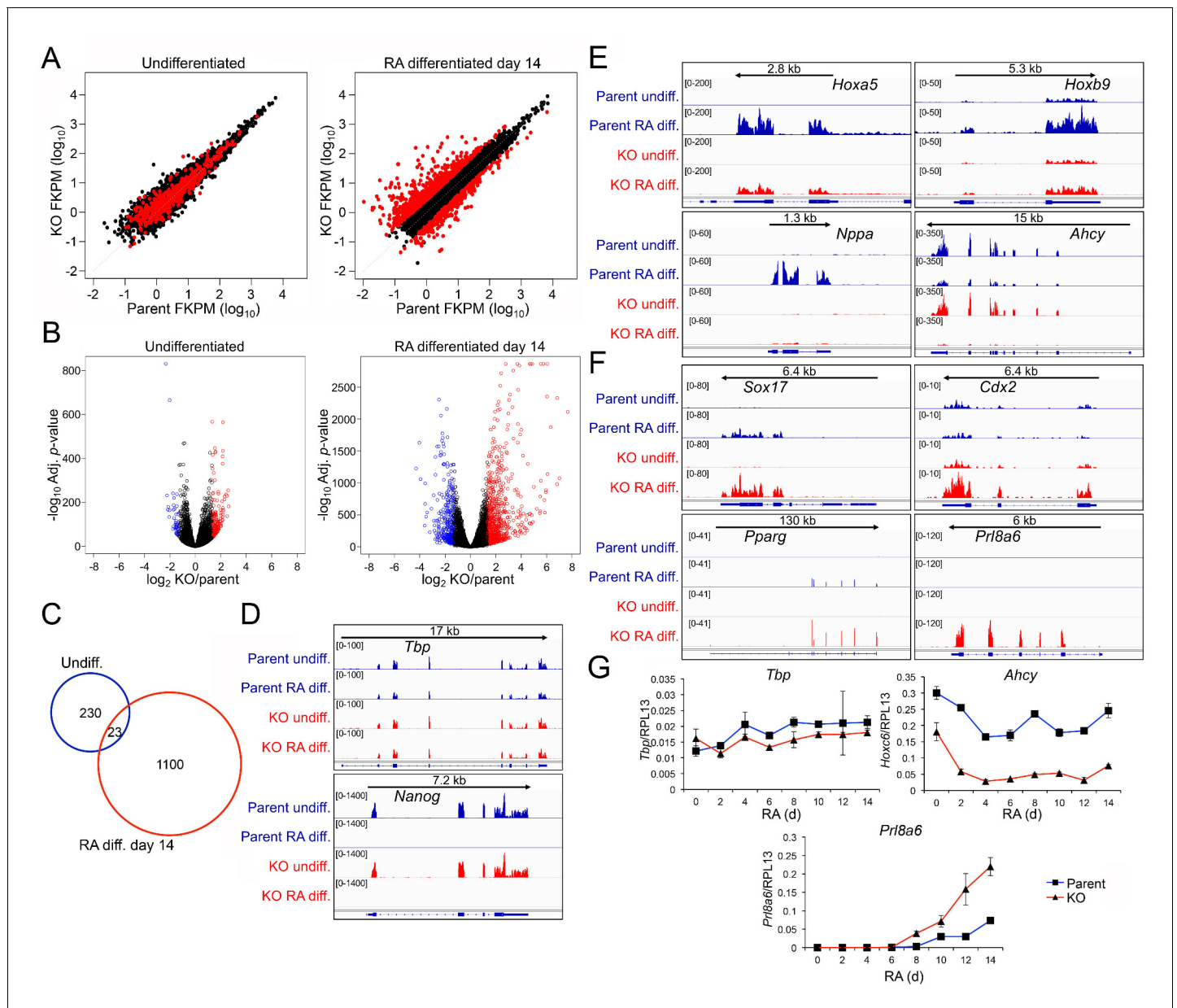
The ability of Oct1 to suppress genes for alternative developmental lineages is reminiscent of findings using T cells in which Oct1 suppresses alternative T cell lineage genes via inter-chromosomal communication between gene loci that execute opposing gene expression programs (**Kim et al., 2014**). Oct1 interacts with CTCF (**Kim et al., 2014**), helping it foster exclusive gene expression programs in T cells. More work is required to determine if Oct1 insures mutually exclusive embryonic developmental gene expression programs through similar mechanisms.

## Materials and methods

### Derivation of Oct1 germline and conditional ESCs

All mice were C57BL/6J background. Oct1 germline-deficient ESCs were generated by intercrossing heterozygous *Pou2f1*<sup>-/+</sup> mice (**Wang et al., 2004**) to generate a 1:2:1 ratio of *Pou2f1*<sup>-/-</sup>: *Pou2f1*<sup>-/+</sup>: *Pou2f1*<sup>+/+</sup> embryonic offspring. ESCs were derived from preimplantation blastocysts and genotyped. Heterozygous ESCs were not studied further. Littermate WT ESCs lines constituted the controls for these experiments. Oct1 inducible-conditional ESCs were generated by first separately crossing mice with the *Pou2f1* conditional (floxed) allele (**Shakya et al., 2015b**) to the YFP reporter B6.129 × 1-Gt(*ROSA*)26Sor<sup>tm1(EYFP)Cos/J</sup> (Jackson labs #006148) and inducible cre transgenic line B6.129-Gt(*ROSA*)26Sor<sup>tm1(cre/ERT2)Tyj/J</sup> (Jackson labs #008463). Resulting *Pou2f1*<sup>fl/fl</sup> animals were intercrossed to generate embryonic *Pou2f1*<sup>fl/fl</sup> offspring in which LSL-YFP was expressed from one *Rosa26* allele and Cre-ERT2 was expressed from the other. Parent ESCs were derived from these preimplantation blastocysts. The parent lines constituted the controls for derived 4-OHT-treated, *Pou2f1*<sup>Δ/Δ</sup>:YFP<sup>+</sup> lines. Cell lines were routinely authenticated by genotyping. Mycoplasma testing was conducted regularly in-house using a previously published method (**Molla Kazemiha et al., 2009**). Cells were negative throughout the study.





**Figure 6.** Genome-wide changes in developmental gene expression following differentiation of Oct1 conditional-inducible deficient ESCs. (A) For each gene, averaged RNAseq FKPM (aligned Fragments Per Kilobase per Million aligned reads) values from three undifferentiated or 14-day RA-differentiated parent and KO ESCs were plotted on a  $\log_{10}$  scale. Red genes signify significantly changed gene expression (Adj.  $p < 0.01$ , fold change  $> 2.5$ ) at d 14. For each timepoint, genes showing  $< 50$  total reads in both genotypes conditions were called unexpressed and are not displayed. (B) Volcano plots showing  $\log_2$  averaged difference in gene expression vs.  $-10 \times \log_{10}$  significance. Significantly altered genes (Adj.  $p < 0.01$ , 2.5-fold change) are shown in blue (down-regulated) or red (up-regulated). (C) Venn diagram showing total numbers of significantly expressed ( $> 2.5$  fold) genes in undifferentiated and 14-day RA-differentiated Oct1-deficient ESCs. Overlap shows genes differentially expressed at both timepoints. (D) Genome tracks of averaged RNAseq read densities (genome build mm10) for two control genes: *Tbp* (a constitutively expressed gene), and *Nanog* (expressed in pluripotent but not differentiated conditions). Arrows show directionality of gene transcription and size of the transcription unit. (E) Additional genome tracks are shown of three genes with poor induction in the KO condition: *Hoxa5*, *Hoxb9*, and *Nppa*. *Ahcy* is also shown, which becomes more strongly down-regulated in the differentiated condition. *Hoxa5* physically overlaps with *Hoxa3*, *Hoxaas3*, and *Hoxa6*, which are not highlighted. (F) Additional genome tracks are shown of genes showing ectopic expression in the differentiated condition: *Sox17*, *Cdx2*, *Pparg*, and *Prl8a6*. (G) RT-qPCR validations of additional genes identified by RNAseq, *Ahcy* and *Prl8a6*. *Tbp* is shown as a control. Average of three biological replicates  $\pm$  standard deviation is shown.

DOI: 10.7554/eLife.20937.025

The following figure supplements are available for figure 6:

Figure 6 continued on next page

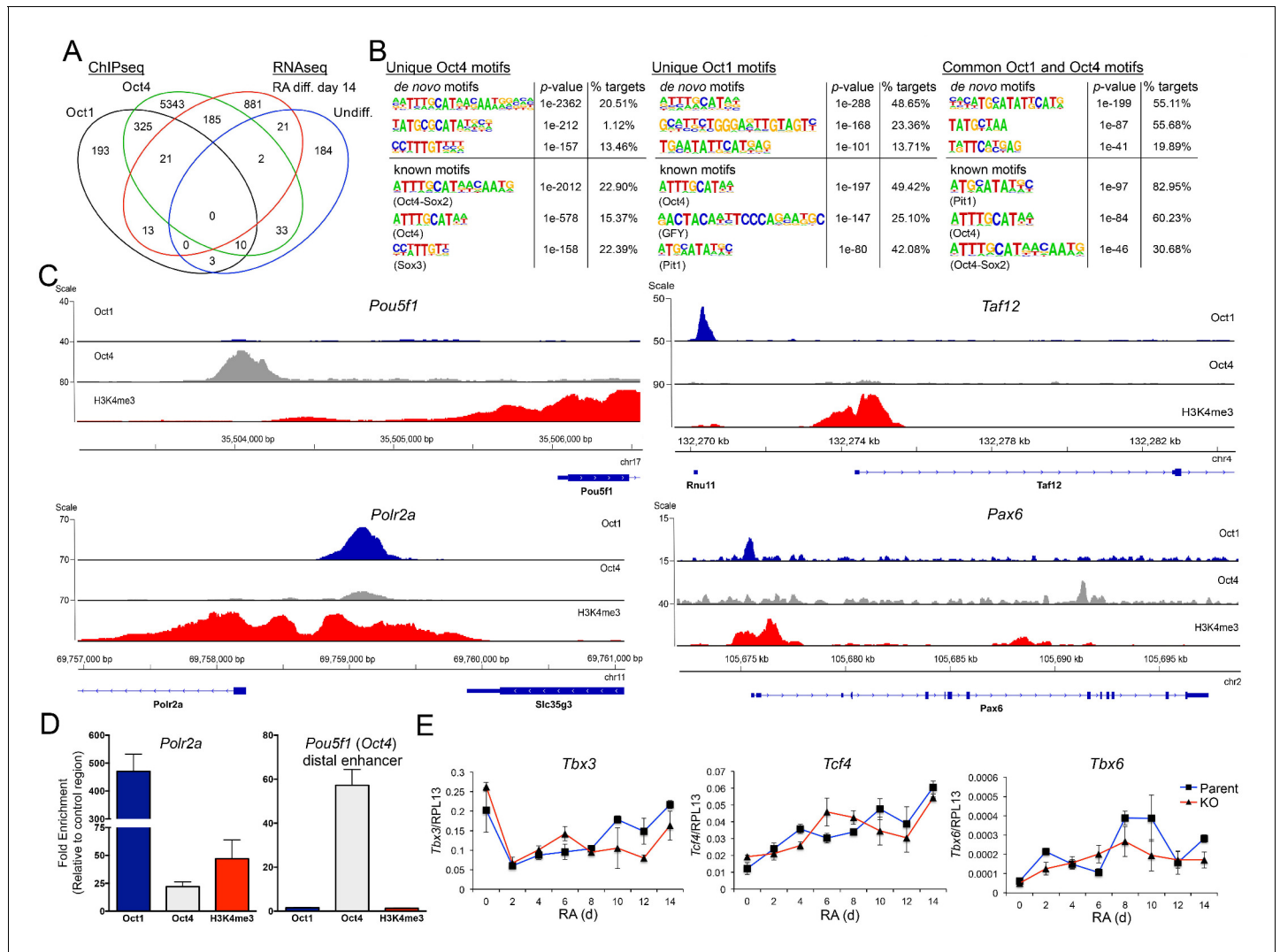
Figure 6 continued

**Figure supplement 1.** Differences in gene expression in differentiated Oct1 deficient cells revealed by RNAseq.

DOI: 10.7554/eLife.20937.026

**Figure supplement 2.** Differences in gene expression in differentiated Oct1 deficient cells revealed by RNAseq.

DOI: 10.7554/eLife.20937.027



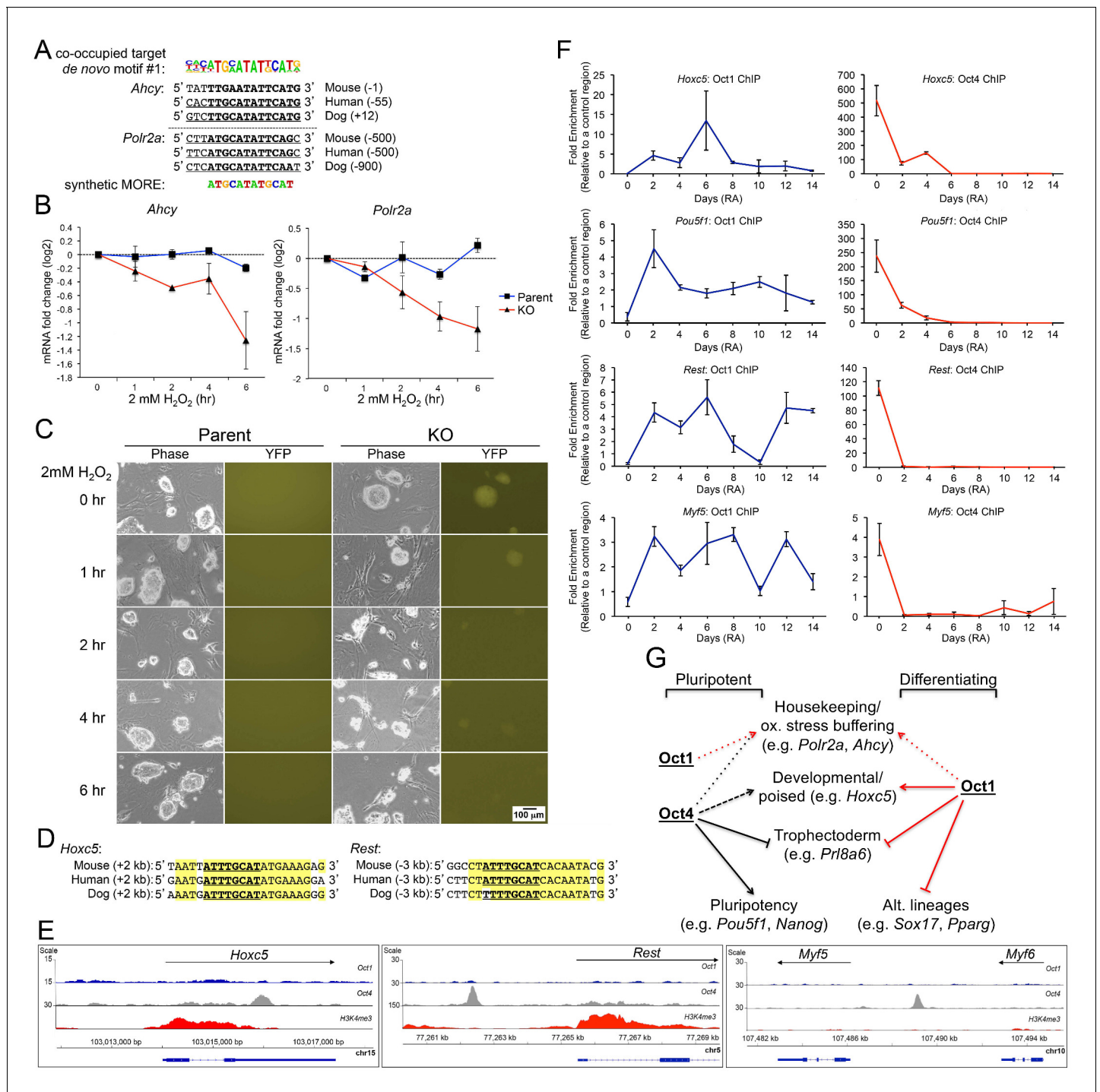
**Figure 7.** Unique and common Oct1 and Oct4 targets in ESCs. (A) Venn diagram illustrating Oct1 and Oct4 target gene profile and intersection with RNAseq gene set. (B) Motif analysis for peaks occupied uniquely by Oct1 or Oct4, and for peaks occupied by both proteins. Top shows best consensus sequences associated with binding. Bottom shows best matches to annotated weight matrices. In the case of known motifs, deviation of physiological binding sites from consensus causes recurring sequences meet threshold criteria for the compound 'Oct4-Sox2' site but not for a simple octamer site ('Oct4'). This is why the percentage of target sites computationally associated with 'Oct4-Sox2' is higher than for 'Oct4.' (C) Genomic tracks showing ChIPseq enrichment for Oct1 (blue), Oct4 (gray) or H3K4me3 (red). Target gene and orientation is shown at the bottom of each track. *Pou5f1*, *Polr2a*, *Taf12* and *Pax6* are shown. (D) ChIP-qPCR validation of select ChIPseq targets. Fold enrichments using Oct1 and Oct4 antibodies at *Polr2a* and *Pou5f1* are shown. (E) RT-qPCR for three identified Oct1 target genes, *Tbx3*, *Tcf4* and *Tbx6* at 0, 2, 4, 6, 8, 10, 12 and 14 d of differentiation. Expression was normalized to the control ribosomal gene RPL13. Three biological replicates were performed. Error bars denote  $\pm$ standard deviation.

DOI: 10.7554/eLife.20937.028

The following figure supplement is available for figure 7:

**Figure supplement 1.** Oct1 and Oct4 ChIPseq read density at example co-bound genes.

DOI: 10.7554/eLife.20937.029



**Figure 8.** Oct1/Oct4 co-binding to MOREs in ESCs, and inducible Oct1 binding to poised targets upon differentiation. (A) Conserved MOREs at two genes (*Ahcy* and *Polr2a*) co-bound by Oct1 and Oct4 in ESCs. The top co-bound de novo motif from **Figure 7B** is shown at top. The MORE sequence (Reményi et al., 2001; Tomilin et al., 2000) is shown at bottom. Mammalian (mouse, human, dog) conservation is shown. MORE position relative to TSS is shown in parentheses. *Polr2a* contains two adjacent MOREs (Kang et al., 2009), only one of which is shown here. (B) Parent and derived KO ESCs were treated with 2 mM H<sub>2</sub>O<sub>2</sub> for the indicated times. *Ahcy* and *Polr2a* log<sub>2</sub> mRNA levels were assessed by RT-qPCR. Three biological replicates were performed. Error bars denote ±standard deviation. (C) Microscopic images of the same cells during the treatment timecourse. (D) Conservation of the octamer sequence in the *Hoxc5* 3' UTR and *Rest/Nrsf* upstream region. Mammalian (mouse, human, dog) conservation is shown. Octamer element position relative to TSS is shown in parentheses. (E) Genome tracks showing ChIPseq enrichment at *Hoxc5*, *Rest* and upstream of *Myf5*, for Oct1 (blue), Oct4 (gray) or H3K4me3 (red). (F) ChIP-qPCR differentiation timecourse of four targets that exclusively bind Oct4 in ESCs. Fold enrichments using Oct1 and Oct4 antibodies at *Hoxc5*, *Pou5f1*, *Rest* and *Myf5* are shown. (G) Model for Oct1 and Oct4 function bind ESCs and their differentiated progeny. The Figure 8 continued on next page

Figure 8 continued

example of neuronal differentiation is shown. In stem cells, Oct1 and Oct4 collaborate at constitutively expressed MORE-containing targets such as *Polr2a* and *Ahcy* to insulate them against oxidative stress (red and black short dashed lines). Oct4 poises developmental genes of all embryonic lineages (long dashed black line) and repress trophoblast-specific genes (solid black block line). Oct4 additionally activates pluripotency genes (solid black arrow). In differentiating cells, Oct1 occupies MORE genes in response to oxidative stress and buffers their expression, as described previously (Kang et al., 2009; Shakya et al., 2011). Oct1 also contributes to eventual lineage-specific developmental gene activation (solid red line), and alternate developmental lineage gene repression (including trophoblast, solid red block line).

DOI: 10.7554/eLife.20937.030

## Cell culture

ESCs were cultured as previously (Shakya et al., 2015a) with 2i conditions: the ERK inhibitor PD0325901 (1  $\mu$ M, LC Laboratories) and the GSK3 inhibitor CHIR99021 (3  $\mu$ M, LC Laboratories). 4-OHT (Sigma) was dissolved in ethanol and used at 500 nM for 24 hr. Two methods were used to generate EBs. Low-attachment dishes were used to generate WT and Oct1-deficient EBs for microscopic analysis, RT-qPCR and the generation of neurons. Briefly, ESCs were trypsinized and feeders depleted by binding to gelatin-coated dishes for 30–60 min. ESC suspensions were plated on low-attachment dishes for 5–7 days. For cardiomyocyte differentiation, the hanging drop method (Wang and Yang, 2008) was used in order to generate single EBs in 96-well plates. Individual EBs were then used to generate cardiomyocyte colonies in 24-well plates. Generation of neurons was accomplished as in (Bain et al., 1995), with modifications. Briefly, EBs were formed for 4 days using low-attachment dishes, followed by culture for a further 4 days as EBs in 0.1  $\mu$ M RA/DMEM. After 8 days, EBs were trypsinized and cultured for 8 days in 1:1 F12:DMEM, 10  $\mu$ g/mL insulin (SAFC Biosciences), 5.5  $\mu$ g/mL transferrin and 38.7  $\mu$ M sodium selenite (ThermoFisher) on laminin/poly-L-lysine-coated ChamberSlides (Corning). Cells in Figure 2E–F were cultured for eight additional d. For H<sub>2</sub>O<sub>2</sub> treatment, ESCs were seeded 24 hr prior to treatment on 6-well plates with sparse feeders. Cells were treated with 2 mM H<sub>2</sub>O<sub>2</sub> (Sigma) for the indicated times.

## Immunoblotting

Antibodies for immunoblotting were as follows: Oct4, Santa Cruz sc-5279; Oct1, Bethyl A301-716A + A301-171A; Nanog, GeneTex GTX100863; Sox2, GeneTex GTX101507; GAPDH, EMD Millipore, MAB374;  $\alpha$ -Tubulin, Santa Cruz sc-5286.

## RT-qPCR

RNA was isolated using TRIzol (Thermo Fisher, Waltham MA), followed by RNAeasy purification (Qiagen) using the RNA cleanup procedure. cDNA was synthesized using SuperScript III and random hexamers (Thermo Fisher). RT-qPCR oligonucleotide primers are listed in **Supplementary file 3**.

## Lentiviral Oct1 complementation

The *Oct1* (*Pou2f1*) cDNA and IRES (internal ribosomal entry site) elements were amplified and cloned together by overlap PCR. In the first PCR, primers to the 5' end of *Oct1* containing a *NotI* restriction site and to the 3' end of *Oct1* that contained a 5' extension of IRES-complementary DNA were used. The sequences were: Oct1-NotI-For: 5'-AATGAAAAAAGCGGCCCGCCATGAATAATCCA TCAGAAAC-3'; Oct1-Rev-IRES: 5'-TTAGGGGGGGGGGAGGGATCTTCACTGTGCCTTGGAG-3'. In the second PCR, an IRES sequence was amplified using primers to the 5' end of the IRES containing a 5' extension of DNA complementary of the *Oct1* 3' end, and primers to the 3' end of the IRES containing an *NdeI* restriction site. The sequences were: IRES-overlap-FOR: 5' AGATCCC TCCCCCCCCCTAACGTTACTGGCCGAA-3'; IRES-Rev-NdeI: 5'-GGGAATCCATATGTGTGGCCA TATTATCATCGTGT-3'. The third PCR used as a template the PCR products from the first two rounds, along with the Oct1-NotI-For and IRES-Rev-NdeI primers. This process generated a DNA fragment containing an *Oct1* cDNA fused to an IRES at the 3' end, along with a *NotI* site at the 5' terminus and an *NdeI* site at the 3' terminus. The fragment was cloned into the optimized, self-inactivating, nonreplicative pHAGE lentiviral vector using the *NotI* and *NdeI* restriction sites. To insert a *Puro* cassette after the IRES, the cDNA was amplified using primers containing 5' *NdeI* and 3' *Clal* restriction sites. The sequences were *Puro*-NdeI-For: 5'-GGAATCCATATGATGACCGAG



TACAAGCCCACGGT-3'; Puro-Clal-Rev: 5' GGTTTATCGATTCAGGCACCGGGCTTGC-3'. Because the IRES apparently attenuated expression of the *Puro* resistance cassette in this vector, puromycin selection was performed at 0.75  $\mu\text{g}/\text{mL}$ . To generate an empty vector control, the vector was cut with *NdeI* and *NotI*, filled in with Klenow fragment, and re-ligated.

## Immunofluorescence

Immunofluorescence was performed as described previously (*Kang et al., 2013*), using mouse monoclonal antibodies against  $\beta$ -tubulin-III (R and D Systems MAB1195) and rabbit polyclonal antibodies against YFP (Life Technologies A6455). Secondary antibodies used were goat anti-rabbit-Alexa568 (Life Technologies A-11011) and goat anti-mouse-Alexa488 (Life Technologies A-11001).

## Teratoma formation

Teratomas were generated as described (*Nelakanti et al., 2015*) by injecting parent or KO ESCs into contralateral flanks of female NCr Nude mice (NCRNU-F, Taconic). Mice were sacrificed at four wk. Tumors were excised, washed with cold PBS and weighed. 1/3 of the excised tumor was used to make lysates for protein analysis using a Dounce homogenizer with RIPA lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, 1 mM EDTA and protease inhibitors [Roche]) on ice. Lysates were centrifuged  $10,000 \times g$  for 10 min. Supernatant protein concentrations were normalized using Bradford assays.  $6 \times$  Laemmli sample buffer was added. The mixture was boiled for 5 min and resolved using a 10% SDS-PAGE gel. The remainder of the tumor was fixed in formaldehyde, paraffin-embedded, sectioned and H and E stained for histological analysis by a blinded pathologist.

## RNAseq

RNA was prepared from three independent cultures of undifferentiated or 14 d RA-differentiated parent *Pou2f1<sup>fl/fl</sup>* or 4-OHT treated *Pou2f1 <sup>$\Delta/\Delta$</sup>*  ESCs. Concentration was determined using a Quant-iT RNA assay kit and a Qubit fluorometer (Thermo Fisher). Intact poly(A) RNA was purified from total RNA samples (100–500 ng) with oligo(dT) magnetic beads, and stranded mRNA sequencing libraries were prepared as described using the Illumina TruSeq mRNA library preparation kit. Purified libraries were qualified on an Agilent Technologies 2200 TapeStation using a D1000 ScreenTape assay. Molarity of adapter-modified molecules was defined by qPCR using the Kapa Biosystems Library Quant Kit. Individual libraries were normalized to 10 nM and equal volumes were pooled in preparation for Illumina sequencing. Sequencing libraries (25 pM) were chemically denatured and applied to an Illumina HiSeq v4 paired end flow cell using an Illumina cBot. Hybridized molecules were clonally amplified and annealed to sequencing primers with reagents from an Illumina HiSeq PE Cluster Kit v4-cBot. Following transfer of the flowcell to an Illumina HiSeq 2500 instrument (HCS v2.2.38 and RTA v1.18.61), a 125-cycle paired-end sequence run was performed using HiSeq SBS Kit v4 sequencing reagents. Fastq data quality were checked using Fastqc version 0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Quality scores dipped towards the 3' end of the reads, so reads were trimmed at 50 bases to eliminate poor-quality data. The resulting 50-base reads were aligned to the mouse mm10 genome (GRCm38, December 2011) plus splice junctions using novoalign version 2.08.01 (<http://www.novocraft.com>). Alignments to splice junctions were translated back to genome coordinates using the SamTranscriptomeParser application in the USeq package (*Nix et al., 2010*). Aligned reads were quality checked using the Picard tools' CollectRnaSeqMetrics command (<https://broadinstitute.github.io/picard/>). On average 99.0% of the reads aligned to the mouse genome, with 78% of reads providing unique alignments, and 86% of reads providing alignments to protein coding and UTR regions of the genome. Tests for differential gene expression were performed with DESeq2, version 1.10.0 (*Love et al., 2014*). Genes with a count of at least 50 in one or more samples were tested. Genes showing at least 2.5-fold change of expression at an adjusted *p*-value of  $<0.01$  were selected as differentially expressed. Figures were generated in R version 3.2.3 (<http://www.r-project.org>) using functions from the gdata and gplots libraries.

## ChIP/ChIPseq

ChIP was performed as described (*Shakya et al., 2015a*). ChIP oligonucleotide primers are listed in **Supplementary file 3**. Antibodies used were the following: Oct1 (Bethyl, a mixture of A301-

716A + A301–717A), Oct4 (Santa Cruz, sc-8629) and H3K4me3 (Millipore, 07–473). ChIPseq was performed as described previously (*Shakya et al., 2015a, 2015b*), using a single IP per condition and clones of parent or derived Oct1-deficient ESCs. For ChIPseq, reads were aligned to the mouse reference genome (mm10) with the Burrows-Wheeler Aligner (BWA, version: 0.5.9). Reads were filtered for alignment quality of >Q10 and duplicates were removed using Picard tools (function MarkDuplicates). After filtering there were 21.1 (H3K4me3), 27.3 (Oct1), and 23.7 (Oct4) million reads. MACSv2 peak caller (version: 2.1) was used to call ChIPseq regions of enrichment with the following parameters (-p 1e-5 -nomodel -shiftsize <fragment\_length/2> for Oct1, Oct4 and -p 1e-2 -broad for H3K4me3). To estimate the -shiftsize parameter (predominant fragment length divided by 2) we performed strand cross-correlation analysis using SPP R package (version: 1.10.1) with default parameters. Peaks overlapping with ENCODE blacklisted regions were filtered using BEDtools (function intersectBed). We also discarded peaks localized to mitochondria, chromosome Y, and unmapped contigs. After filtering we had 692 (Oct1), and 8673 (Oct4) peaks. Signal to noise ratio was assessed by calculating normalized strand coefficient (NSC) and relative strand correlation (RSC) using the SPP R package with default parameters (version: 1.10.1). The obtained values of NSC and RSC (H3K4me3: 2.28, 1.25; Oct1: 1.02, 1.45; Oct4: 1.05, 2.32) indicate highly enriched datasets with large fragment-length peak as compared to read-length peak. The NSC value for Oct1 transcription factor was somewhat smaller but typical for high quality datasets generated for factors with small numbers of genuine binding sites (692 MACS2-identified peaks for Oct1). We used MACSv2 function bdgdiff to build fold-enrichment signal tracks for all positions in the genome. Signal tracks were converted to TDF files using igvtools (<https://www.broadinstitute.org/igv/igvtools>). Peaks were allocated to genes using the annotatePeaks.pl program from HOMER suite (Hypergeometric Optimization of Motif Enrichment, version: 4.7, <http://homer.salk.edu/homer/>) by determining the closest RefSeq transcription start sites of the genes to the peaks. Functional enrichment analysis was performed using the findGO.pl program from HOMER and Bonferroni as well as Benjamini and Hochberg correction for multiple testing corrections. Robustness of the analysis was confirmed using MEME-ChIP (*Machanic and Bailey, 2011*), which generate highly similar motifs.

## Motif analysis

Transcription factor enrichment within ChIPseq peaks (de novo motif discovery and known motif matching) was determined using findMotifsGenome.pl program from HOMER. Motif analysis was run on overlapped and separately on unique Oct1 and Oct4 ChIPseq peaks. Oct1 and Oct4 ChIPseq peak overlaps were defined by requiring the distance between peak summits to be  $\leq 100$  bp. Motif lengths of 6–24 bp were identified within 200 bp regions centered on peak summits and an option of random background was selected for motif discovery.

## Acknowledgements

We thank Drs. Time Formosa and David Stillman for critical reading of the manuscript. We thank Susan Tamowski and the University of Utah Health Sciences Center Transgenic and Gene Targeting Core for assistance generating Oct1 deficient and conditional-inducible ESCs, and for assistance with blastocyst injections to generate chimeras. We thank James Cox and the University of Utah Metabolomics Core. We thank Brian Dalley and the University of Utah High-Throughput Genomics Core, Brett Milash and the University of Utah Bioinformatics Core, and the University of Utah Center for High Performance Computing for assistance with RNAseq. This work was supported by grants from the University of Utah Department of Pathology, NIH/NIAID (R01AI100873) and an endowment (Watkins Endowed Chair) to DT

---

## Additional information

### Competing interests

AR: Senior editor, *eLife*. The other authors declare that no competing interests exist.

**Funding**

Funder	Grant reference number	Author
National Institute of Allergy and Infectious Diseases	R01AI100873	Dean Tantin

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

**Author contributions**

ZS, JK, AS, MT, Data curation, Formal analysis, Investigation, Writing—original draft, Writing—review and editing; EAJ, Formal analysis, Writing—original draft, Writing—review and editing; AR, Supervision, Funding acquisition, Writing—original draft, Project administration, Writing—review and editing; DT, Conceptualization, Supervision, Funding acquisition, Writing—original draft, Project administration, Writing—review and editing

**Author ORCIDs**

Dean Tantin,  <http://orcid.org/0000-0003-1354-8385>

**Ethics**

Animal experimentation: This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All of the animals were handled according to approved institutional animal care and use committee (IACUC) protocols (#14-06015) of the University of Utah. Every effort was made to minimize suffering.

**Additional files****Supplementary files**

- Supplementary file 1. Normalized RNAseq gene expression changes.  
DOI: [10.7554/eLife.20937.031](https://doi.org/10.7554/eLife.20937.031)
- Supplementary file 2. Genes identified by ChIPseq in comparison to RNAseq data.  
DOI: [10.7554/eLife.20937.032](https://doi.org/10.7554/eLife.20937.032)
- Supplementary file 3. Oligonucleotides for RT-qPCR and ChIP.  
DOI: [10.7554/eLife.20937.033](https://doi.org/10.7554/eLife.20937.033)

**Major datasets**

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Shen Z, Kang J, Shakya A, Tabaka M, Jarboe EA, Regev A, Tantin D	2017	Enforcement of developmental lineage specificity by transcription factor Oct1	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85063">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85063</a>	Publicly available at the NCBI Gene Expression Omnibus (accession no: GSE85063)

**References**

- Abranches E**, Silva M, Pradier L, Schulz H, Hummel O, Henrique D, Bekman E. 2009. Neural differentiation of embryonic stem cells in vitro: a road map to neurogenesis in the embryo. *PLoS One* **4**:e6286. doi: [10.1371/journal.pone.0006286](https://doi.org/10.1371/journal.pone.0006286), PMID: [19621087](https://pubmed.ncbi.nlm.nih.gov/19621087/)
- Azuara V**, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG. 2006. Chromatin signatures of pluripotent cell lines. *Nature Cell Biology* **8**:532–538. doi: [10.1038/ncb1403](https://doi.org/10.1038/ncb1403), PMID: [16570078](https://pubmed.ncbi.nlm.nih.gov/16570078/)
- Bain G**, Kitchens D, Yao M, Huettner JE, Gottlieb DI. 1995. Embryonic stem cells express neuronal properties in vitro. *Developmental Biology* **168**:342–357. doi: [10.1006/dbio.1995.1085](https://doi.org/10.1006/dbio.1995.1085), PMID: [7729574](https://pubmed.ncbi.nlm.nih.gov/7729574/)
- Bellance N**, Pabst L, Allen G, Rossignol R, Nagrath D. 2012. Oncosecretomics coupled to bioenergetics identifies  $\alpha$ -amino adipic acid, isoleucine and GABA as potential biomarkers of Cancer: differential expression of c-Myc,



- Oct1 and KLF4 coordinates metabolic changes. *Biochimica Et Biophysica Acta (BBA) - Bioenergetics* **1817**: 2060–2071. doi: [10.1016/j.bbabi.2012.07.004](https://doi.org/10.1016/j.bbabi.2012.07.004), PMID: 22842522
- Bernstein BE**, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**:315–326. doi: [10.1016/j.cell.2006.02.041](https://doi.org/10.1016/j.cell.2006.02.041), PMID: 16630819
- Boyer LA**, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**:947–956. doi: [10.1016/j.cell.2005.08.020](https://doi.org/10.1016/j.cell.2005.08.020), PMID: 16153702
- Chen X**, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**:1106–1117. doi: [10.1016/j.cell.2008.04.043](https://doi.org/10.1016/j.cell.2008.04.043), PMID: 18555785
- Chen J**, Liu H, Liu J, Qi J, Wei B, Yang J, Liang H, Chen Y, Chen J, Wu Y, Guo L, Zhu J, Zhao X, Peng T, Zhang Y, Chen S, Li X, Li D, Wang T, Pei D. 2013. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genetics* **45**:34–42. doi: [10.1038/ng.2491](https://doi.org/10.1038/ng.2491), PMID: 23202127
- De Los Angeles A**, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, Hochedlinger K, Jaenisch R, Lee S, Leitch HG, Lensch MW, Lujan E, Pei D, Rossant J, Wernig M, Park PJ, Daley GO. 2015. Hallmarks of pluripotency. *Nature* **525**:469–478. doi: [10.1038/nature15515](https://doi.org/10.1038/nature15515), PMID: 26399828
- Guo RJ**, Suh ER, Lynch JP. 2004. The role of cdx proteins in intestinal development and Cancer. *Cancer Biology & Therapy* **3**:593–601. doi: [10.4161/cbt.3.7.913](https://doi.org/10.4161/cbt.3.7.913), PMID: 15136761
- Jiang H**, Shukla A, Wang X, Chen WY, Bernstein BE, Roeder RG. 2011. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* **144**:513–525. doi: [10.1016/j.cell.2011.01.020](https://doi.org/10.1016/j.cell.2011.01.020), PMID: 21335234
- Jin T**, Li H. 2001. Pou homeodomain protein OCT1 is implicated in the expression of the caudal-related homeobox gene Cdx-2. *Journal of Biological Chemistry* **276**:14752–14758. doi: [10.1074/jbc.M008277200](https://doi.org/10.1074/jbc.M008277200), PMID: 11278400
- Kang J**, Gemberling M, Nakamura M, Whitby FG, Handa H, Fairbrother WG, Tantin D. 2009. A general mechanism for transcription regulation by Oct1 and Oct4 in response to genotoxic and oxidative stress. *Genes & Development* **23**:208–222. doi: [10.1101/gad.1750709](https://doi.org/10.1101/gad.1750709), PMID: 19171782
- Kang J**, Shen Z, Lim JM, Handa H, Wells L, Tantin D. 2013. Regulation of Oct1/Pou2f1 transcription activity by O-GlcNAcylation. *The FASEB Journal* **27**:2807–2817. doi: [10.1096/fj.12-220897](https://doi.org/10.1096/fj.12-220897), PMID: 23580612
- Kim LK**, Esplugues E, Zorca CE, Parisi F, Kluger Y, Kim TH, Galjart NJ, Flavell RA. 2014. Oct-1 regulates IL-17 expression by directing interchromosomal associations in conjunction with CTCF in T cells. *Molecular Cell* **54**: 56–66. doi: [10.1016/j.molcel.2014.02.004](https://doi.org/10.1016/j.molcel.2014.02.004)
- Ku M**, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genetics* **4**:e1000242. doi: [10.1371/journal.pgen.1000242](https://doi.org/10.1371/journal.pgen.1000242), PMID: 18974828
- Love MI**, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. doi: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8), PMID: 25516281
- Lu CW**, Yabuuchi A, Chen L, Viswanathan S, Kim K, Daley GO. 2008. Ras-MAPK signaling promotes trophoblast formation from embryonic stem cells and mouse embryos. *Nature Genetics* **40**:921–926. doi: [10.1038/ng.173](https://doi.org/10.1038/ng.173), PMID: 18536715
- Machanic P**, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**:1696–1697. doi: [10.1093/bioinformatics/btr189](https://doi.org/10.1093/bioinformatics/btr189), PMID: 21486936
- Maddox J**, Shakya A, South S, Shelton D, Andersen JN, Chidester S, Kang J, Gligorich KM, Jones DA, Spangrude GJ, Welm BE, Tantin D. 2012. Transcription factor Oct1 is a somatic and Cancer stem cell determinant. *PLoS Genetics* **8**:e1003048. doi: [10.1371/journal.pgen.1003048](https://doi.org/10.1371/journal.pgen.1003048), PMID: 23144633
- Meissner A**, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**:766–770. doi: [10.1038/nature07107](https://doi.org/10.1038/nature07107), PMID: 18600261
- Molla Kazemihah V**, Shokrgozar MA, Arabestani MR, Shojaei Moghadam M, Azari S, Maleki S, Amanzadeh A, Jeddi Tehrani M, Shokri F. 2009. PCR-based detection and eradication of mycoplasma infections from various mammalian cell lines: a local experience. *Cytotechnology* **61**:117–124. doi: [10.1007/s10616-010-9252-6](https://doi.org/10.1007/s10616-010-9252-6), PMID: 20135349
- Morey L**, Santanach A, Di Croce L. 2015. Pluripotency and Epigenetic factors in mouse embryonic stem cell Fate Regulation. *Molecular and Cellular Biology* **35**:2716–2728. doi: [10.1128/MCB.00266-15](https://doi.org/10.1128/MCB.00266-15), PMID: 26031336
- Nelakanti RV**, Kooreman NG, Wu JC. 2015. Teratoma formation: a tool for monitoring pluripotency in stem cell research. *Current Protocols in Stem Cell Biology* **32**:4A.8.1–4.8.4. doi: [10.1002/9780470151808.sc04a08s32](https://doi.org/10.1002/9780470151808.sc04a08s32), PMID: 25640819
- Nichols J**, Zevnik B, Anastasiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Schöler H, Smith A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**:379–391. doi: [10.1016/S0092-8674\(00\)81769-9](https://doi.org/10.1016/S0092-8674(00)81769-9), PMID: 9814708
- Nix DA**, Di Sera TL, Dalley BK, Milash BA, Cundick RM, Quinn KS, Courdy SJ. 2010. Next generation tools for genomic data generation, distribution, and visualization. *BMC Bioinformatics* **11**:455. doi: [10.1186/1471-2105-11-455](https://doi.org/10.1186/1471-2105-11-455), PMID: 20828407

- Okamoto K**, Okazawa H, Okuda A, Sakai M, Muramatsu M, Hamada H. 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**:461–472. doi: [10.1016/0092-8674\(90\)90597-8](https://doi.org/10.1016/0092-8674(90)90597-8), PMID: [1967980](https://pubmed.ncbi.nlm.nih.gov/1967980/)
- Pan G**, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**:299–312. doi: [10.1016/j.stem.2007.08.003](https://doi.org/10.1016/j.stem.2007.08.003), PMID: [18371364](https://pubmed.ncbi.nlm.nih.gov/18371364/)
- Reményi A**, Tomilin A, Pohl E, Lins K, Philippsen A, Reinbold R, Schöler HR, Wilmanns M. 2001. Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Molecular Cell* **8**:569–580. doi: [10.1016/S1097-2765\(01\)00336-7](https://doi.org/10.1016/S1097-2765(01)00336-7), PMID: [11583619](https://pubmed.ncbi.nlm.nih.gov/11583619/)
- Rosner MH**, Vigano MA, Ozato K, Timmons PM, Poirier F, Rigby PW, Staudt LM. 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**:686–692. doi: [10.1038/345686a0](https://doi.org/10.1038/345686a0), PMID: [1972777](https://pubmed.ncbi.nlm.nih.gov/1972777/)
- Sebastiano V**, Dalvai M, Gentile L, Schubart K, Sutter J, Wu GM, Tapia N, Esch D, Ju JY, Hübner K, Bravo MJ, Schöler HR, Cavaleri F, Matthias P. 2010. Oct1 regulates trophoblast development during early mouse embryogenesis. *Development* **137**:3551–3560. doi: [10.1242/dev.047027](https://doi.org/10.1242/dev.047027), PMID: [20876643](https://pubmed.ncbi.nlm.nih.gov/20876643/)
- Shakya A**, Cooksey R, Cox JE, Wang V, McClain DA, Tantin D. 2009. Oct1 loss of function induces a coordinate metabolic shift that opposes tumorigenicity. *Nature Cell Biology* **11**:320–327. doi: [10.1038/ncb1840](https://doi.org/10.1038/ncb1840), PMID: [19219035](https://pubmed.ncbi.nlm.nih.gov/19219035/)
- Shakya A**, Kang J, Chumley J, Williams MA, Tantin D. 2011. Oct1 is a switchable, bipotential stabilizer of repressed and inducible transcriptional states. *Journal of Biological Chemistry* **286**:450–459. doi: [10.1074/jbc.M110.174045](https://doi.org/10.1074/jbc.M110.174045), PMID: [21051540](https://pubmed.ncbi.nlm.nih.gov/21051540/)
- Shakya A**, Callister C, Goren A, Yosef N, Garg N, Khoddami V, Nix D, Regev A, Tantin D. 2015a. Pluripotency transcription factor Oct4 mediates stepwise nucleosome demethylation and depletion. *Molecular and Cellular Biology* **35**:1014–1025. doi: [10.1128/MCB.01105-14](https://doi.org/10.1128/MCB.01105-14), PMID: [25582194](https://pubmed.ncbi.nlm.nih.gov/25582194/)
- Shakya A**, Goren A, Shalek A, German CN, Snook J, Kuchroo VK, Yosef N, Chan RC, Regev A, Williams MA, Tantin D. 2015b. Oct1 and OCA-B are selectively required for CD4 memory T cell function. *The Journal of Experimental Medicine* **212**:2115–2131. doi: [10.1084/jem.20150363](https://doi.org/10.1084/jem.20150363), PMID: [26481684](https://pubmed.ncbi.nlm.nih.gov/26481684/)
- Sridharan R**, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M, Carey M, Garcia BA, Plath K. 2013. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1 $\gamma$  in reprogramming to pluripotency. *Nature Cell Biology* **15**:872–882. doi: [10.1038/ncb2768](https://doi.org/10.1038/ncb2768), PMID: [23748610](https://pubmed.ncbi.nlm.nih.gov/23748610/)
- Suh E**, Traber PG. 1996. An intestine-specific homeobox gene regulates proliferation and differentiation. *Molecular and Cellular Biology* **16**:619–625. doi: [10.1128/MCB.16.2.619](https://doi.org/10.1128/MCB.16.2.619), PMID: [8552090](https://pubmed.ncbi.nlm.nih.gov/8552090/)
- Takahashi K**, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**:663–676. doi: [10.1016/j.cell.2006.07.024](https://doi.org/10.1016/j.cell.2006.07.024), PMID: [16904174](https://pubmed.ncbi.nlm.nih.gov/16904174/)
- Tantin D**, Schild-Poulter C, Wang V, Haché RJ, Sharp PA. 2005. The octamer binding transcription factor Oct-1 is a stress sensor. *Cancer Research* **65**:10750–10758. doi: [10.1158/0008-5472.CAN-05-2399](https://doi.org/10.1158/0008-5472.CAN-05-2399), PMID: [16322220](https://pubmed.ncbi.nlm.nih.gov/16322220/)
- Tantin D**, Gemberling M, Callister C, Fairbrother WG, Fairbrother W. 2008. High-throughput biochemical analysis of in vivo location data reveals novel distinct classes of POU5F1(Oct4)/DNA complexes. *Genome Research* **18**:631–639. doi: [10.1101/gr.072942.107](https://doi.org/10.1101/gr.072942.107), PMID: [18212089](https://pubmed.ncbi.nlm.nih.gov/18212089/)
- Tantin D**. 2013. Oct transcription factors in development and stem cells: insights and mechanisms. *Development* **140**:2857–2866. doi: [10.1242/dev.095927](https://doi.org/10.1242/dev.095927), PMID: [23821033](https://pubmed.ncbi.nlm.nih.gov/23821033/)
- Theodorou E**, Dalembert G, Heffelfinger C, White E, Weissman S, Corcoran L, Snyder M. 2009. A high throughput embryonic stem cell screen identifies Oct-2 as a bifunctional regulator of neuronal differentiation. *Genes & Development* **23**:575–588. doi: [10.1101/gad.1772509](https://doi.org/10.1101/gad.1772509), PMID: [19270158](https://pubmed.ncbi.nlm.nih.gov/19270158/)
- Tomilin A**, Reményi A, Lins K, Bak H, Leidel S, Vriend G, Wilmanns M, Schöler HR. 2000. Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell* **103**:853–864. doi: [10.1016/S0092-8674\(00\)00189-6](https://doi.org/10.1016/S0092-8674(00)00189-6), PMID: [11136971](https://pubmed.ncbi.nlm.nih.gov/11136971/)
- Verrijzer CP**, van Oosterhout JA, van der Vliet PC. 1992. The Oct-1 POU domain mediates interactions between Oct-1 and other POU proteins. *Molecular and Cellular Biology* **12**:542–551. doi: [10.1128/MCB.12.2.542](https://doi.org/10.1128/MCB.12.2.542), PMID: [1346336](https://pubmed.ncbi.nlm.nih.gov/1346336/)
- Vázquez-Arreguín K**, Tantin D. 2016. The Oct1 transcription factor and epithelial malignancies: old protein learns new tricks. *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1859**:792–804. doi: [10.1016/j.bbagr.2016.02.007](https://doi.org/10.1016/j.bbagr.2016.02.007), PMID: [26877236](https://pubmed.ncbi.nlm.nih.gov/26877236/)
- Walker E**, Ohishi M, Davey RE, Zhang W, Cassar PA, Tanaka TS, Der SD, Morris Q, Hughes TR, Zandstra PW, Stanford WL. 2007. Prediction and testing of novel transcriptional networks regulating embryonic stem cell self-renewal and commitment. *Cell Stem Cell* **1**:71–86. doi: [10.1016/j.stem.2007.04.002](https://doi.org/10.1016/j.stem.2007.04.002), PMID: [18371337](https://pubmed.ncbi.nlm.nih.gov/18371337/)
- Wang VE**, Schmidt T, Chen J, Sharp PA, Tantin D. 2004. Embryonic lethality, decreased erythropoiesis, and defective octamer-dependent promoter activation in Oct-1-deficient mice. *Molecular and Cellular Biology* **24**:1022–1032. doi: [10.1128/MCB.24.3.1022-1032.2004](https://doi.org/10.1128/MCB.24.3.1022-1032.2004), PMID: [14729950](https://pubmed.ncbi.nlm.nih.gov/14729950/)
- Wang X**, Yang P. 2008. In vitro differentiation of mouse embryonic stem (mES) cells using the hanging drop method. *Journal of Visualized Experiments* **23**:825. doi: [10.3791/825](https://doi.org/10.3791/825), PMID: [19066514](https://pubmed.ncbi.nlm.nih.gov/19066514/)
- Wang Y**, Yabuuchi A, McKinney-Freeman S, Ducharme DM, Ray MK, Chawengsaksohak K, Archer TK, Daley GQ. 2008. Cdx gene deficiency compromises embryonic hematopoiesis in the mouse. *PNAS* **105**:7756–7761. doi: [10.1073/pnas.0708951105](https://doi.org/10.1073/pnas.0708951105), PMID: [18511567](https://pubmed.ncbi.nlm.nih.gov/18511567/)

- Wang P**, Wang Q, Sun J, Wu J, Li H, Zhang N, Huang Y, Su B, Li RK, Liu L, Zhang Y, Elsholtz HP, Hu J, Gaisano HY, Jin T. 2009. POU homeodomain protein Oct-1 functions as a sensor for cyclic AMP. *Journal of Biological Chemistry* **284**:26456–26465. doi: [10.1074/jbc.M109.030668](https://doi.org/10.1074/jbc.M109.030668), PMID: [19617623](https://pubmed.ncbi.nlm.nih.gov/19617623/)
- Wen B**, Wu H, Shinkai Y, Irizarry RA, Feinberg AP. 2009. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genetics* **41**:246–250. doi: [10.1038/ng.297](https://doi.org/10.1038/ng.297), PMID: [19151716](https://pubmed.ncbi.nlm.nih.gov/19151716/)
- Yeap LS**, Hayashi K, Surani MA. 2009. ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. *Epigenetics & Chromatin* **2**:12. doi: [10.1186/1756-8935-2-12](https://doi.org/10.1186/1756-8935-2-12), PMID: [19811652](https://pubmed.ncbi.nlm.nih.gov/19811652/)
- Yuan P**, Han J, Guo G, Orlov YL, Huss M, Loh YH, Yaw LP, Robson P, Lim B, Ng HH. 2009. Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes & Development* **23**:2507–2520. doi: [10.1101/gad.1831909](https://doi.org/10.1101/gad.1831909), PMID: [19884257](https://pubmed.ncbi.nlm.nih.gov/19884257/)
- Zylicz JJ**, Dietmann S, Günesdogan U, Hackett JA, Cougot D, Lee C, Surani MA. 2015. Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. *eLife* **4**:09571. doi: [10.7554/eLife.09571](https://doi.org/10.7554/eLife.09571), PMID: [26551560](https://pubmed.ncbi.nlm.nih.gov/26551560/)