# Oct1 recruits the histone lysine demethylase Utx to canalize lineage specification

**Jelena Perovanovic[1,3], Yifan Wu[1,3], Zuolian Shen[1,3], Erik Hughes[1,3], Mahesh B. Chandrasekharan[2,3], Dean Tantin[1,3,*]**

[1]Department of Pathology

[2]Department of Radiation Oncology

[3]Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84112 U.S.A.

*Correspondence

dean.tantin@path.utah.edu

Short title: Oct1 mediates gene bivalency resolution

Key words: pre-implantation development; embryonic stem cells; single-cell RNA sequencing; Oct1/Pou2f1; Utx/Kmd3a

Characters in title (including spaces): 82

Words in abstract: 150

Character count (including spaces and main legends but excluding STAR methods, Supplemental item legends and References): 66,316

Total main figures and tables: 7

## HIGHLIGHTS

- Oct1 "canalizes" gene expression to promote developmentally accurate gene expression and differentiation
- Oct1 is necessary for proper bivalency resolution
- Oct1 recruits Utx to demethylate H3K27me3 at mesoderm lineage-appropriate targets

## SUMMARY

**The pathways used by cells to transition between undifferentiated, pluripotent state and tissue-specific states are incompletely understood. Here we show that the transcription factor Oct1/Pou2f1 activates silent, developmental lineage-appropriate genes to "canalize" developmental progression. Using inducible knockout embryonic stem cells and single-cell gene expression profiling, we show that Oct1 deficiency impairs mesodermal differentiation. Oct1-deficient cells show inappropriate developmental lineage branching and display a poorly differentiated state with hallmarks of pluripotency. Like Oct4, Oct1 directly binds genes critical for mesoderm induction and chromatin modification. Oct1 recruits the Utx/Kdm3a histone lysine demethylase to remove inhibitory H3K27me3 marks and activate their expression. The specificity of the ubiquitous Oct1 protein for mesodermal genes is explained by cooperative interactions with the Smad3 transcription factor. We also show that ectopic Oct1 expression improves the ability of cells to differentiate accurately under mesoderm lineage-inducing conditions. Overall, these results identify Oct1 as a key regulator of mesoderm differentiation.**

## INTRODUCTION

Lineage specification is a key process in the development of multicellular organisms by which undifferentiated cells progressively acquire tissue- and cell type-specific features (Seydoux and Braun, 2006). It is dynamically regulated, requiring extensive transcriptional and epigenetic remodeling to selectively activate lineage-appropriate gene expression programs and stably repress the expression of genes specific for alternative lineages. Embryonic stem cells (ESCs) represent a pluripotent cell type capable of both self-renewal and differentiation into all three germ layers (Beddington and Robertson, 1989). The three germ layers are established during gastrulation – a spatial reorganization of the embryo from a single-layer epiblast into a multilayered organism. One of the germ layers – mesoderm (MD) – gives rise to dermomyotome (muscle and dermis), sclerotome (axial skeleton) and lateral MD (cardiac) among other tissue types.

The transcriptional changes underlying lineage specification require extensive chromatin remodeling and spatiotemporal activation of genes encoding master transcription factors. Remodeling is potentiated by a unique chromatin landscape in pluripotent cells. Chromatin in ESCs is largely accessible and lacks stable heterochromatin domains (Meshorer and Misteli, 2006; Schlesinger and Meshorer, 2019). A large number of genes encoding lineage-specific developmental regulators are marked at promoters and gene bodies by covalently modified nucleosomes that simultaneously convey activating (H3K4me3) and repressing (H3K27me3) potential (Bernstein et al., 2006; Ku et al., 2008). In ESCs, these “bivalent” genes are silent or expressed at low levels, but poised for activation. During development, gene bivalency resolves via either removal of activating marks and gene silencing, or removal of repressive marks and gene activation. Which bivalent genes resolve in which direction is lineage-specific, resulting in distinct, cell fate-appropriate transcription programs and durable repression of lineage-inappropriate genes. However, the programs that instruct correct bivalency resolution are poorly understood.

POU transcription factors play central roles in the regulation of development (Tantin, 2013). The well-known POU factor Oct4 (encoded by *Pou5f1*) is a master regulator of the induction and maintenance of pluripotency (Nichols et al., 1998; Palmieri et al., 1994; Takahashi and Yamanaka, 2006). Oct4 associates with bivalent genes in ESCs (Bernstein et al., 2006), but is rapidly silenced in their differentiated progeny before bivalency resolution and the induction of tissue- and cell type-specific gene expression. A second POU protein, Oct1, is co-expressed with

Oct4, but unlike Oct4 is expressed beyond pluripotency (Shen et al., 2017). Oct1 is widely expressed and required for placental and embryonic development (Sebastiano et al., 2010; Wang et al., 2004). Circumventing the placental defects via tetraploid complementation results in developmental arrest at E8.25 with no more than five somites (Sebastiano et al., 2010). Oct1-deficient ESCs are morphologically normal and show appropriate self-renewal and growth characteristics, but upon differentiation show phenotypic and gene expression defects including decreased developmental lineage-specific gene expression coupled with elevated expression of developmentally incorrect genes (Shen et al., 2017). The underlying molecular mechanisms by which Oct1 regulates lineage differentiation are unknown.

Here, using MD differentiation of *Oct1*/*Pou2f1* inducible-conditional ESCs, we show that Oct1 "canalizes" lineage specification by directly de-repressing lineage-specific bivalent genes. Oct1 is needed for consequent early MD development and adoption of terminal MD phenotypes. In the absence of Oct1, differentiating cells show "fuzzy" differentiation programs and increased predilection to proceed down incorrect developmental trajectories. We show that Oct1 interacts with components of the Utx (Kdm3a) H3K27me3 demethylase complex, and recruits Utx to developmentally appropriate target genes to facilitate bivalency resolution. Binding sites for Oct1 and Oct4 in pluripotent cells in which Oct1 binding is carried forward during MD differentiation are also enriched in sites for Smad transcription factors, key mediators of MD specification. Oct1 and Smad3 interact in differentiating cells, providing a means of restricting Oct1's bivalency-resolving potential to MD-specific genes. Finally, we show that ectopic Oct1 expression increases MD developmental specification, suggesting an application in improving the specificity of developmental outcomes in differentiating ESCs or iPSCs.

## RESULTS

### Loss of Oct1 Causes Aberrant Mesodermal Differentiation

To identify how Oct1 loss perturbs cells during differentiation, we performed single-cell RNA-seq (scRNA-seq) using an MD differentiation protocol that results in terminally-differentiated myotubes after 3 weeks of culture. The differentiation process is initiated by Bmp4 treatment to activate TGFβ signaling, and generates dynamic changes in metabolic programs (Oginuma et al., 2017), epithelial-to-mesenchymal transition (EMT) (Diaz-Cuadros et al., 2020) and induction of MD-specific gene expression programs during the differentiation timecourse (Chal et al., 2015). Cells were collected at day 0 (D0, pluripotent ESCs), D3 and D6 of differentiation to identify changes in populations and gene expression patterns early in the differentiation timecourse. Data from the different conditions were subjected to graph-based clustering approach and uniform manifold approximation and projection (UMAP) to visualize clusters of cells with similar gene expression. Integrated analysis of Oct1-sufficient *Pou2f1*$^{fl/fl}$;Rosa26-CreER$^{T2}$;LSL-YFP conditional ESCs (hereafter, "parental" cells) at the three timepoints identifies a range of cell populations, from ESC-like to neuromesodermal, paraxial MD and early somite (Figure 1A). These populations recapitulate developmental stages associated with induction of markers associated with early MD (*Fgf17* and *Tbxt*/*T*/*Brachyury*; cluster 2 and 7), neuromesoderm (*Hoxc9*, *Pax3*, *Sox2*), paraxial MD (*Meox1*, *Twist1*, *Pax3*) and early somite and neuronal progenitors (*Cxcr4, Uncx*) (Figure 1B, Figure S1and not shown). These results provide a single cell-resolution map of early MD differentiation.

Next, we compared the parental controls to tamoxifen-treated, Oct1-deficient cells (hereafter, "cKO") to identify changes in cell populations and gene expression (Table S1). D0 and D3 cKO cells show few differences from parental, Oct1 sufficient cells (not shown). In contrast, D6 cKO cells show poor mesodermal differentiation capacity with reductions in key mesodermal clusters. For example, parental cell populations characterized by neuromosodermal and paraxial MD gene expression represent the largest two clusters, accounting for 14.7 and 12.9% of cells, respectively, while a population associated with somites accounts for 5.6% of cells (Figure 1C left panel). These findings are consistent with findings that somites are derived from multiple transcriptional trajectories including early paraxial mesoderm and neuromesodermal progenitors (Guibentif et al., 2021). In contrast, cKO D6 cells show increases in cells that retain epithelial characteristics and dramatically decreased percentages of neuromesodermal progeitors (2.3%), paraxial MD (7.5%), and somites (0.5%, Figure 1C, right panel). Comparing gene expression

between parental and cKO paraxial mesoderm clusters, cKO cells fail to appropriately induce lineage-appropriate genes such as *Pax3* and *Pax7*, and inappropriately upregulate lineage-inappropriate markers such as the epithelial-specific genes *Krt8* and *Krt18* (Figure 1D). These results show that Oct1 is necessary for accurate mesoderm differentiation and to suppress expression of genes for alternative lineages.

Pseudotime analysis of scRNA-seq data allows multiple differentiation timepoints to be overlaid with defined starting (ESC) and endpoint (somite) populations. Parental control cells progress through a largely linear pathway with one minor branch retaining an inappropriate epithelial fate (Figure 2A, top panel). In contrast, differentiating cKO ESCs show a larger proportion inappropriate branching into alternative developmental trajectories (bottom panel), consistent with the diminished developmental progression of paraxial MD to the somite stage, and consistent with enrichment of cells that inappropriately maintain an epithelial state (Figure 1C and Figure S1). We also examined pseudotemporal gene expression using specific genes associated with pluripotency and MD development. In this analysis, the position of each cell in pseudotime is shown on the X-axis and gene expression on the Y-axis (Figure 2B). Parental cells show robust early expression of genes associated with pluripotency such *as Klf4* that lose expression roughly mid-way through pseudotime. *Tbxt* is transiently induced, while the early somite markers *Pax7* and *Cxcr4* are efficiently and coordinately induced later (Figure 2B, top panel). In contrast, cKO cells exhibit inappropriately prolonged expression of *Klf4* and *Tbxt*, and largely fail to upregulate *Pax7* and *Cxcr4* (bottom panel). We used pseudotemporal ordering to visualize expression of the 2000 most variable (dynamic) genes in parental cells, and compared expression of the same genes in the same order in cKO cells (Figure 2C). This analysis revealed clusters associated with pluripotency (*Klf4*, *Nanog*), early MD differentiation (*T*), paraxial mesoderm (*Rspo3, Meox1*), and early somites (*Dll1, Pax3*). The pseudotime gene expression pattern for cKO cells revealed clusters with “fuzzy” patterns of gene expression early in differentiation, including prolonged expression of genes such as *Tbxt* and poor and blurred expression of genes in the clusters associated with paraxial mesoderm. In addition, the induction of somitic genes normally expressed later in development such as *Dll1*, *Pax3* and *Pax7* was poor in cKO cells (Figure 2C).

RNA velocity analysis allows developmental directionality to be inferred based on the ratio of unspliced, newly-synthesized pre-mRNA to mature spliced mRNA. A vector is then assigned to each cell that indicates developmental direction and magnitude relative to the other cells. Applying these algorithms, we found that D6 differentiated parental cells form discrete clusters

with dynamic developmental progression, e.g. paraxial MD and somite (Figure 2D, marked by long arrows). cKO cells by contrast are marked by stationary profiles indicative of failed differentiation potential, with cells progressing primarily towards a poorly differentiated state characterized by multiple lineage markers (Figure 2D). Cumulatively, the data demonstrate that Oct1 is required for efficient progression of differentiation program activation, with Oct1-deficient cells unable to canalize appropriate lineage specification programs.

**Oct1 Occupies Developmental Genes in ESCs**

One model that explains the above data is that Oct1 occupies developmental-specific targets bound by Oct4 in ESCs, to mediate induction of developmentally appropriate genes and repression of genes specific to alternative lineages. To test this hypothesis, we performed Oct1 ChIP-seq. Prior work showed that Oct1 occupancy at Oct4 targets increases during retinoic acid (RA)-mediated differentiation when Oct4 expression is lost (Shen et al., 2017). RA ultimately differentiates cells into neuroectodermal lineages (Bain et al., 1995). In undifferentiated ESCs, strong Oct1 binding (with Oct4) was only observed in a group of ~100 genes containing Oct protein variant binding sites termed MOREs (Shen et al., 2017). We used a different Oct1 antibody with superior enrichment properties (see methods) to perform ChIP-seq with undifferentiated parental Oct1-sufficient ESCs, as well as cells differentiated towards MD for 3 and 6 days. Oct4 ChIP-seq in undifferentiated ESCs was performed as a parallel control.

In pluripotent cells, ~22,000 Oct4 peaks were identified, corresponding to ~6,000 genes with transcription start sites (TSS) within 20 kb (Table S2). ~45% of Oct4 targets directly overlap Oct1 peaks. Conversely ~60% of Oct1 targets overlap Oct4 peaks (Figure 3A). The shared targets in ESCs are on average more strongly enriched for Oct4 than Oct1 as shown by tag density (Figure 3B), although the different antibodies could also contribute to the differential enrichment. Shared Oct1/Oct4 targets in ESCs include *Polr2a*, which encodes the largest subunit of RNA polymerase II, *Pou5f1*, which encodes Oct4, and *Dll1*, which encodes a developmentally-inducible mediator of Notch signaling expressed in the MD lineage where it regulates muscle development (Zhang et al., 2021). Tracks are shown in Figure S2A. We confirmed Oct1 and Oct4 binding to *Polr2a*, *Pou5f1* and *Dll1* using ChIP-qPCR (Figure S2). Oct1 binding to *Polr2a*, which contains two adjacent MOREs that can bind four Oct proteins (Kang et al., 2009a), is exceptional in that it is far stronger than Oct4 relative to other genes (100× stronger for *Polr2a*, 3-10× weaker for *Pou5f1* and *Dll*). Re-ChIP (sequential ChIP) indicates that Oct1 and Oct4 bind these sites simultaneously. The signal was specific to Oct1 because it was lost in Oct1 cKO ESCs (Figure

S2). Cumulatively, the data indicate that in ESCs Oct1 co-binds with Oct4 to an array of targets, including developmental-specific targets.

We performed ChIP-seq using D3 and D6 MD-differentiated cells. Only ~200 Oct4-bound peaks not occupied by Oct1 in pluripotent cells become occupied by Oct1 at D6 of MD differentiation (Figure 3A). Another ~800 peaks shared by Oct4 and Oct1 in pluripotent cells continue to be bound by Oct1 at D6. Analysis of these peaks using GREAT (McLean et al., 2010) identifies enrichment for oxidative stress, ribosomal and mitochondrial organization, Notch signaling and post-implantation development including somite formation (Figure S3A). Also, >6000 peaks become uniquely bound by Oct1 at MD D6 (Figure 3A). ~2300 peaks are uniquely bound at D3 (Figure S3B).

To pattern Oct1 and Oct4 occupancy during MD differentiation, we applied hierarchical clustering. Three major gene clusters were identified, three of which show static or decreased Oct1 binding over time (Figure 3C, clusters 1, 2, 4). Gene ontogeny (GO) analysis indicates involvement in signaling, mitochondrial function, stem cell differentiation and the regulation of Notch and TGFβ signaling. In contrast, Oct1 binding increases with differentiation in cluster 3. Genes near these peaks are associated with chromatin-modifying activities including acetyltransferase and polycomb group complexes. Oct1 and Oct4 enrichment at example cluster 4 genes encoding polycomb complex members is shown in Figure 3D. These genes (*Ezh2*, *Suz12*, *Ring1* and *Ezh1*) show a gain in Oct1 occupancy as cells progress from pluripotency through D3 to D6 of MD differentiation. In the scRNA-seq UMAP projection, *Ezh2* expression is downregulated in cKO cells at D6, in particular in paraxial mesoderm (Figure 3E, F). The pseudotime induction of *Ezh2* expression observed in parental cells also fails in cKO (Figure 3G). These data indicate that during MD differentiation Oct1 directly binds to, and mediates the induction of, genes encoding epigenetic regulators.

We also queried Oct1 binding to genes associated with bivalency. To identify bivalent genes, we intersected ESC H3K27me3- and H3K4me3-enriched ChIP-seq peaks from the ENCODE database. This procedure identifies 3861 bivalent genes (Table S3). To query Oct1 occupancy on these genes during differentiation, we intersected the bivalent gene dataset with Oct1 ChIP-seq in pluripotent and MD-differentiated parental cells, observing an increase in binding over time (Figure 3H, green and blue lines). A similar analysis at D6 using just MD-specific bivalent genes (generated by intersecting the bivalent gene dataset with MD development GO:0007498, Table S4) shows even stronger Oct1 enrichment (Figure 3H, purple line). These

findings indicate that Oct1 robustly binds to bivalent genes both in pluripotent cells and their differentiated progeny, with binding increasing at lineage-specific genes duriung differentiation.

**Oct1 is Necessary for Appropriate Bivalency Resolution**

To investigate pertubations in gene expression caused by Oct1 loss, we performed RNA-seq using parental and cKO ESCs, and cells differentiated towards MD for 3 or 6 days. Three replicates were performed per condition. ~1700, ~800 and 3,300 significantly differentially expressed genes were identified at D0, D3 and D6, respectively (FDR $\leq$ 0.05; -1 < log2FC < 1, Table S5). Euclidean distance analysis reveals tight correlations between replicates and conditions at D0 and D3 (Figure 4A), but strong divergence at D6 relative to the other conditions and between parental and cKO D6 replicates (Figure 4A). Unsupervised hierarchical clustering reveals groups of genes regulated similarly between parental and cKO cells, and others with differential expression. One group, cluster 2, was marked by strong induction in differentiating cKO but failed induction in parental D6 cells. Example genes in this set include *Pax3*, *Pax7* and *Myog* (Figure 4B, box). GO analysis of the differentially expressed genes in cluster 2 reveals association with differentiation to neural tube (neuroectoderm) and somite (mesoderm, Figure 4C). By contrast, clusters 1 and 4, which were associated with mis-expression of genes in differentiated cKO cells, were associated with terms such as primordial germ cells, immune system, vascular system and stromal cells (not shown). To identify potential regulators of these genes, we queried the ChIP Enrichment Analysis (ChEA) database (Lachmann et al., 2010). Genes with failed induction in differentiating cKO cells tend to be bound by the polycomb repressor compex 2 (PRC2) components Suz12 and Mtf2 (Figure 4C). Example tracks are shown in Figure 4D for *Pou5f1* (pluripotency), *Tbxt* (*T*/*Brachyury*, cluster 3) and *Pax7* (cluster 2). The retained D3 expression of *Pou5f1*, which encodes Oct4, provides a likely explanation for the tight correlations between gene expression states at D0 and D3, as these cells maintain pluripotency characteristics through D3. These data indicate that Oct1 loss results in defective induction of developmental genes that are also regulated by chromatin-modifying activities that act on H3K27me3.

**Oct1 directly regulates H3K27me3 via the recruitment of Utx**

Bivalency is positively resolved via the removal of repressive H3K27me3 marks to activate lineage-appropriate genes, or negatively resolved via the removal of activating H3K4me3 marks to stably repress lineage-inappropriate genes (Bernstein et al., 2006; Dhar et al., 2016). Our

findings indicate that genes with failed induction in differentiating cKO cells tend to be bound by chromatin-modifying complexes that act on H3K27me3. We hypothesized that Oct1 locally recruits H3K27-specific demethylases to lineage-specific bivalent genes to mediate their induction during differentiation. One such demethylase is Utx/Kdm6a (Cloos et al., 2008), which has been shown to regulate neuronal differentiation. To test the association between Oct1 binding, H3K27me3 and Utx, we also performed ChIP-seq using specific antibodies at D6 of differentiation of parental, Oct1 sufficient cells. ~12,000 H3K27me3 peaks and ~12,000 Utx peaks were identified, corresponding to ~11,000 genes within 20 kb (Table S2). Unsupervised hierarchical clustering together with Oct1 peaks from the same cells identifies shared and distinct peaks, including weak Oct1 binding events with strong Utx and H3K27me3 association (cluster 3), and strong Oct1 binding events associated with weaker Utx and lack of H3K27me3 (cluster 4, Figure 5A). This latter cluster includes genes induced in MD such as *Pax3, Pax7* and *Myog*. We interpret cluster 3 to be poised bivalent genes and cluster 4 to be positively resolved genes. GO terms associated with these clusters were enriched for development and musculoskeletal abnormalities (Figure 5B). Intersecting the Utx and Oct1 ChIP-seq peaks identifies a high degree of overlap, with ~70% of Oct1-bound peaks also associating with Utx (Figure 5C).

To test if cKO cells inappropriately retain H3K27me3 at lineage-appropriate genes during MD differentiation, we performed ChIP-qPCR using primers flanking a strong Oct1 peak on the *Pax3* promoter. *Pax3* is induced by D6 of MD differentiation (Figure 2C), and reduced in Oct1 cKO cells at this timepoint (Figure 2C and Figure 1D). As expected, H3K27me3 is robust and equivalent in D0 undifferentiated cells, while D6 parental cells show reduced H3K27me3. In contrast, cKO cells inappropriately retain elevated H3K27me3 (Figure 5D). Consequently, ~3-fold higher H3K27me3 enrichment is observed in differentiated cKO relative to parental cells. Global H3K27me3 levels are unchanged at the same timepoint, as evidenced by immunofluorescence (Figure 5E). Next, we performed Utx ChIP-qPCR with the same primers. Utx is enriched at the *Pax3* promoter at D6 of MD differentiation, validating the ChIP-seq enrichment in parental cells. Further, Utx showed reduced enrichment in cKO cells (Figure 5F). To demonstrate that Oct1 and Utx were simultaneously bound, we performed re-ChIP experiments, observing enrichment at *Pax3* and another region, an enhancer region near *Dll1*, but not at the 28S control region (Figure 5G). These results indicate simultaneous association of Utx and Oct1 at *Pax3*, and specific defects in removing H3K27me3 from *Pax3* in differentiating Oct1-deficient cells.

We then determined if Oct1-Utx interact by co-immunoprecipitation using extracts from undifferentiated parental ESCs, as well as cells at D6 of MD differentiation. The Oct1-Utx

interaction is established in ESCs and maintained during differentiation (Figure 5H). Utx recruitment by Oct1 potentially explains the failure of bivalent, lineage-specific genes such as *Pax3* to positively resolve bivalency in cKO cells (Figure 5D). Oct1, Oct4 and Utx binding to *Pax3* is shown in Figure 5I.

**Smad proteins cooperate with Oct1 in cells differentiating towards MD**

The broad expression of Oct1 raised the question of how Oct1 recruits Utx specifically at lineage-specific genes to mediate gene activation. Chromatin accessibility, lineage-specific signaling and co-bound transcription regulators may provide this specificity. We performed motif analysis (Heinz et al., 2010) using DNA sequences ±100 bp from the center of the 807 peaks co-bound by Oct1 and Oct4 in pluripotent cells that remain Oct1-bound during MD differentiation (Figure 3A). This procedure identifies not only Oct1 sites, but also binding sites for Smads, the terminal transcription regulators of TGFβ/Nodal signaling (Figure 6A). TGFβ signals and downstream Smad transcription factor activity are critical for MD induction (Conlon et al., 1994; Zhou et al., 1993). A study identified Oct1 transcription factor motifs near binding events for zebrafish Smad2/3 (Liu et al., 2011). This study further showed that ectopic expression of the zebrafish Oct1 ortholog enhances mesoderm induction, that zebrafish Smad2 and Oct1 physically interact, and that the two factors cooperate to enhance transcription. Mammalian Smad3 also interacts with Oct4 and co-occupies target genes in pluripotent cells (Mullen et al., 2011). Consistent with these findings, Oct1 immunoprecipitation in D6 MD-differentiated cells efficienty detects co-precipitated Smad3 (Figure 6B). *Smad3* is expressed in undifferentiated populations, but further induced by MD differentiation (Figure 6C).

To test if Oct1 and Smad proteins such as Smad3 synergize to regulate gene expression, we cloned an 85 bp region ~3 kb upstream of the *Myog* gene together with the CMV core promoter upstream of a luciferase reporter vector (Figure 6D). This region contains two octamer sites and two Smad sites, and overlaps with an ENCODE distal enhancer signature (Abascal et al., 2020). The region is also located within a known super-enhancer associated with myotubes (Hnisz et al., 2013). The sequences were inserted into a reporter vector that expresses secreted nanoLuciferase, and co-transfected with a plasmid constitutively expressing secreted mCherry (Wider and Picard, 2017) as an internal standard (Figure 6D). Cells were either co-transfected with a construct encoding mouse Oct1, treated with recombinant TGFb1, or both. The two treatments together generate cooperative luciferase activity (Figure 6E). No such activity was

observed using constructs with mutant Oct/Smad binding sites. These results indicate that Oct and Smad sites at the *Myog* enhancer cooperate to drive gene expression.

**Loss of Oct1 results in abnormal terminal MD differentiation**

To study the consequences of Oct1 deficiency for MD development, we differentiated parental and cKO cells for 19 days, and queried expression of the myogenic genes *Myod* and *Myog* by RT-qPCR. Relative to Rps2 (which encodes a ribosomal 40S subunit), both genes are strongly expressed in parental but not cKO cells (Figure 7A). Consistently, after D11 of differentiation, immunostaining for embryonic myosin heavy chain (MyH-emb) reveals robust expression in fused myotubes in parental cells, with cKO expression close to background (Figure 7B). These results demonstrate that cKO ESCs differentiate into muscle poorly, consistent with their defective early gene expression programs and developmental trajectories.

Metabolic changes during development are necessary to complete the MD differentiation program (Cliff et al., 2017; Oginuma et al., 2017). Among these changes are the induction of high glycolytic flux during formation of paraxial MD, which is required for posterior elongation of embryonic axis (Oginuma et al., 2017). To determine changes in steady-state metabolites in parental and cKO ESCs at day 0 and 6 of MD differentiation, we performed GC-MS. Principle component analysis of the metabolites showed few differences in metabolite levels (Figure S4A). Differentiation of parental cells significantly changes metabolite composition, in particular in the PC1 axis (Figure S4A). Metabolites markedly changed by differentiation include the amino acids leucine and threonine (up), and TCA intermediates such as malate and fumarate (down, Figure S4B). Compared to parental cells, cKO cells show a significant downregulation ($P<0.007$) of both D-glucose and pyruvate, and an increase ($P<0.01$) in TCA intermediates such as citrate, isocitrate, malate and succinate (Figure S5C). These results suggest that cKO cells induce glycolysis poorly during MD differentiation.

**Ectopic Oct1 expression improves mesodermal specification and differentiation**

To complement Oct1 deficiency and determine the effect of ectopic Oct1 expression in Oct1-deficient cells, we transduced parental and cKO ESCs with retroviral vectors encoding murine Oct1 or empty vector (EV) controls. The vectors encode a puromycin resistance cassette, allowing for selection of transduced cells. Immediately after selection, populations of transduced cells were subjected to MD differentiation for 11 d, and probed for *Myod*, *Myog* and *Pax3* by RT-qPCR. Strikingly, differentiating Oct1-transduced ESCs more strongly express both *Myod* and *Myog*, and

370 decreased *Pax3* (Figure 7C). The combination of elevated *Myod* and *Myog* with decreased *Pax3*
371 expression at late differentiation timepoints suggests that ectopic, retrovirally expressed Oct1 not
372 subject to endogenous regulation enables cells to more efficiently transit through a *Pax3*-
373 expressing intermediate state, such that more cells enter into a terminal myogenic program.
374 Immunoblotting confirmed ectopic Oct1 expression and complementation of Oct1 deficiency
375 (Figure 7D).

## DISCUSSION

During development, chromatin transitions from a pluripotent state permissive for different lineages to a lineage-restricted state. Pluripotent cells maintain genes encoding developmental-specific mediators in a poised "bivalent" configuration that allows for later induction or stable repression, depending on the developmental lineage (Bernstein et al., 2006). Here we show that during mesodermal differentiation, the Oct4 paralog Oct1 binds and mediates the induction of developmental lineage-specific bivalent genes by recruiting the Utx histone demethylase to remove repressive H3K27me3 marks. Consequently, Oct1-deficient ESCs manifest defective MD differentiation, including failure to express *Myod* and *Myog* mRNA, and myosin heavy chain protein. Oct1-deficient animals manifest defective induction of somites, cardiac tissue and blood cells (Sebastiano et al., 2010; Wang et al., 2004).

The central role of Oct1 in "canalizing" differentiating pluripotent cells in early steps of mesodermal specification was shown using single-cell RNA-seq with Oct1 cKO ESCs. In the absence of Oct1, cells undergoing differentiation retain pluripotency characteristics such as an epithelial state, and achieve somite-stage gene expression patterns poorly and in reduced numbers. The cells mis-express developmentally inappropriate genes and undergo inappropriate developmental branching towards poorly differentiated states marked by epithelial gene expression and oxidative stress. The induction of genes important early in MD differentiation such as *Tbxt* becomes weaker and loses temporal coherence. Later, there is failed induction of genes such as *Pax7* and *Cxcr4*.

In differentiating cells, Oct1 occupies a subset of genes bound by Oct4 in pluripotent cells. These include MD-specific bivalent developmental mediators and genes encoding chromatin-modifying enzymes. Lineage-specific bivalent, developmentally poised genes are poorly induced in cKO cells. Oct1 occupancy on these genes normally increases with differentiation, suggesting a critical role in canalization of cell fate. We find that Oct1 recruits Utx to lineage-specific, Oct1-bound targets such as *Pax3*. The lack of Utx recruitment to lineage-specific genes in cKO cells is consistent with the abnormal retention of H3K27me3 at the promoters of these genes and their failed transcriptional upregulation. An "anti-repression" mechanism for Oct1 involving the removal of repressive chromatin marks has been described for Oct1 before, for H3K9me2 (Shakya et al., 2011). This is the first description of a role for Oct1 and K3K27me3. One mechanism that allows Oct1 to use Utx as a cofactor specifically at MD-appropriate targets is collaboration with Smad3. Smad transcription factor binding sites are enriched near sites of Oct1 binding during MD

409 differentiation. TGFβ signals drive Smad transcription factor binding and are critical for MD
410 specification (Liu et al., 2011; Mullen et al., 2011). Co-IP experiments in D6 MD-differentiated
411 cells show an interaction between Oct1 and Smad3, consistent with prior findings in zebrafish (Liu
412 et al., 2011). A model for Oct1's function at lineage-specific genes during MD specification and
413 later differentiation is shown in Figure 7E. In this model, Oct1 and Smad form cooperative
414 complexes at MD-specific genes, and Utx recruitment to Oct1 allows for loss of the repressive
415 H3K27me3 mark and successful resolution of bivalency. Subsequently, other transcription factors
416 (e.g., MyoD) act as primary "on" switches to induce gene expression. Oct1 also binds and induces
417 the expression of genes encoding PRC complex members such as *Ezh2* in MD-specific clusters
418 in parental but not cKO cells. The increased expression of PRC components may solidify lineage
419 specification by aiding the repression of the large number of genes specific to alternative lineages
420 (Collinson et al., 2016).

421 We show that ectopic Oct1 expression can improve canalization and lineage-specific gene
422 expression. Under MD differentiation conditions, exogenous Oct1 increases expression of the
423 terminal differentiation markers *Myod* and *Myog*, while decreasing the early lineage-specification
424 marker *Pax3*, which is transiently expressed during MD differentiation and is necessary for later
425 expression of myogenic genes. Because *Pax3* is no longer expressed in terminally differentiating
426 cells, these results suggest that ectopic Oct1 enables transit through a *Pax3*-expressing
427 intermediate to potentiate productive terminal differentiation. More investigation into this pathway
428 may uncover methods to more efficiently differentiate pluripotent cells. Understanding how to
429 improve canalization of differentiating cells could lead to improvements in therapeutic approaches
430 for regenerative medicine.

## FIGURE LEGENDS

**Figure 1. Differentiating Oct1-deficient ESCs canalize poorly into mesodermal lineages**

(A) UMAP projection of scRNA-seq data from superimposed parental (Oct1 sufficient) undifferentiated ESCs, and parental cells early during MD differentiation (days 3 and 6). Clusters of cells were labeled based the expression of developmental markers as in (B). Three combined replicate plates were used for the analysis.

(B) Violin plots showing gene expression levels of key developmental markers by cluster. Data were log-normalized for each cell using the natural logarithm, scaled and averaged using mean(expm1(x)).

(C) Comparative UMAP projections of integrated D6 parental and Oct1-deficient (cKO) scRNA-seq populations. Clusters were labeled computationally and identified based on gene expression. Relative frequencies are shown.

(D) Differential gene expression analysis of the neuromesodermal cluster shown as a scatter plot. Red dots depict significantly differentially expressed genes based on FDR corrected $p < 0.05$ and fold change >1.2. Example differentially expressed genes are labeled.

**Figure 2. Oct1-deficient cells show perturbed developmental trajectories**

(A) Pseudotime analysis of pluripotent and differentiating parental (top panel) and cKO (bottom panel) cells. Colors correspond to the time point at which cells were collected (red: D0, green: D3, blue: D6).

(B) *Klf4*, *Tbxt*, *Pax7* and *Cxrc4* mRNA expression across pseudotime in parental (top panels) and cKO (bottom panels). Black trendline represents an average expression for a given gene across all populations.

(C) Heatmap depicting expression of the 2000 most dynamically expressed genes (based on FindVariableFeatures function, Seurat) in parental D6 cells. Gene expression was plotted as a heat map across pseudotime in parental (left panel) and cKO (right panel). Dynamic genes were first hierarchically clustered in parental cells to cluster groups of genes that behave similarly in pseudotime, then plotted in the same order in cKO cells.

(D) Velocity gene expression analysis of parental (left panel) and cKO (right panel) cells at differentiation day 6. Arrows point toward cells with gene expression closest to the future state of each cell based on unspliced/spliced transcripts. Arrow length represents magnitude.

**Figure 3. Oct1 co-occupies target sites with Oct4 in ESCs, and regulates their expression during differentiation**

(A) Venn diagram showing common and unique Oct4 and Oct1 binding sites based on ChIP-seq in parental ESCs and at D3 and D6 of mesodermal differentiation.

(B) Tag density around peak centers for total Oct4 (red) and Oct1 (blue) ChIP-seq data sets.

(C) A matrix of Oct4 and Oct1 ChIP-seq enrichment 2 kb around peak centers was computed for the merged peak list and shown as heatmap. Positions in color show high enrichment and white shows no enrichment.

(D) Tracks of representative loci associated with regulation of H3K27me3. Y-axes were scaled to the same value for each gene.

(E) *Ezh2* expression is shown in UMAP projections for parental and cKO cells at D6 of MD differentiation. The paraxial MD cluster is outlined in red.

(F) Violin plots showing *Ezh2* expression in cells within the paraxial MD cluster shown in (E).

(G) *Ezh2* expression in pseudotime in parental (top panel) and cKO (bottom panel) cells. Black trendline represents average expression across all the cells in pseudotime.

(H) Oct1 enrichment based on tag density at peak center at annotated bivalent genes in ESCs (red), and at MD differentiation D3 and D6 (green and blue). An additional analysis was performed for MD-specific bivalent genes at MD differentiation D6 (purple).

**Figure 4. Failed bivalent gene induction in differentiating Oct1 cKO ESCs**

(A) Clustering of the RNAseq replicates using Euclidean distance method shows divergence between replicates and conditions. D6 shows the most variance compared to the other times and comparing parental (WT) and cKO (KO).

(B) Bulk RNA-seq Heatmap of differentially expressed genes (parental vs. cKO) at D0, D3 and D6 of MD differentiation is shown. Cluster 2 shows poor gene induction in the Oct1-deficient condition. Representative genes in this cluster are shown at right.

(C) Jensen Tissue and ChIP-X Enrichment Analysis (ChEA) query results for Cluster 2 in (B) are shown.

(D) Example RNA-seq genome tracks (*Pou5f1*, *Tbxt*, *Pax7*) are shown. *Pax7* is an example cluster 2 gene. Y-axes were group-autoscaled for each gene.

**Figure 5. Oct1 interacts with Utx to demethylate H3K27me3 at bivalent genes.**

(A) A matrix of Oct1, H3K27me3 and Utx ChIP-seq enrichment 2 kb around peak centers was computed for the merged peak list and shown as heatmap. Positions in color show high enrichment and positions in white show no enrichment.

(B) Human Phenotype Ontology terms for Cluster 3 and 4 genes from (A) are shown.

(C) The Oct1 and Utx peak lists were intersected (overlap ≥ 1 bp) and plotted as a Venn diagram. 69% of Oct1 peaks overlapped with Utx binding events.

(D) H3K27me3 ChIP-qPCR enrichment at the *Pax3* promoter in parental and cKO cells at MD differentiation D0 (ESC) and D6. Normalized fold-enrichment is shown relative to both an isotype control antibody and to a nonspecific genomic region encoding the 28S ribosomal subunit. Data represent an average of 3 independent biological replicates. Error bars depict ±SEM.

(E) H3K27me3 immunofluorescence images from D6 MD-differentiated parental and cKO cultures. Images were collected a 40× magnification.

(F) Utx enrichment at the *Pax3* promoter in parental and cKO cells on differentiation D6. Fold-enrichment relative to an isotype control antibody and relative to a 28S ribosomal subunit region is shown. Data represent an average of 3 independent biological replicates. Error bars depict ±SEM.

(G) Sequential ChIP (re-ChIP)-qPCR Oct1/Utx simultaneous enrichment at the *Pax3* promoter and *Dll1* enhancer is shown for both parental and Oct1 cKO cells at D6 of MD differentiation. Fold enrichment relative to an isotype control antibody is shown. 28S is shown as a control. Data represent an average of 3 independent biological replicates. Error bars depict ±SEM.

(H) Immunoprecipitation of Utx followed by Oct1 immunoblot using parental ESCs, or ESCs MD-differentiated for 6D.

(I) Signal tracks (*Mm10* v.D191020) showing Oct4, Oct1 and Utx enrichment at the *Pax3* locus 5' region. Y-axes were group-autoscaled for each gene. Positions of identified HOMER Smad2/3/4 motifs are shown below.

**Figure 6. Oct1 and Smad3 cooperate to drive expression of mesoderm-specific bivalent genes**

(A) HOMER motif analysis of Oct1 peaks that are both shared with Oct4 in ESCs, and maintained after D6 of MD differentiation.

(B) Smad3 immunoblot using cell lysates immunoprecipitated with Oct1 antibodies, or rabbit IgG controls. D6 MD-differentiated cells were used. 20% input is shown (lane 1).

(C) *Smad3* gene expression is shown in violin plots for parental ESCs at D6 of MD differentiation.
(D) Schema for Oct1/Smad reporter assay. A segment of a mouse *Myog* enhancer element containing multiple octamer and Smad motifs was cloned with the core CMV promoter upstream of secreted nLuc and co-transfected into Oct1-deficient MEFs together with a construct encoding constitutive secreted mCherry as a normalization control. Added TGFb1 and co-transfected mouse Oct1 supply Oct1 and Smad3 activity.
(E) Transfected WT (left panel) or Oct/Smad mutant (right panel) *Myog* enhancer constructs were supplied with Oct1, recombinant purified TGFb1 treatment, or both. For each construct, secreted luciferase activity was assessed relative to secreted mCherry expressed from a co-transfected plasmid. An average of three experimental replicates is shown. Error bars denote ±SEM. For the situation in which both Oct1 and TGFb1 are both supplied, fold changes relative to a double-mutant construct are also shown.
(F) Signal tracks (*Mm10* v. D191020) showing Oct1ChIP-seq enrichment at the *Myog* locus. Shown above are RNA-seq tracks in differentiated and undifferentiated parental and cKO cells. Annotated enhancer elements are shown below.

**Figure 7. Myogenic lineage induction is defective in the absence of Oct1 and augmented with Oct1 ectopic expression**

(A) Relative mRNA expression levels of the myogenic genes *Myod1* and *Myog* in parental and cKO ESCs differentiated for 19D. Data represent an average of 3 independent biological replicates. Error bars depict ±SEM.
(B) Embryonic myosin heavy chain (MyHC-emb) expression alone and merged with DAPI is shown using parental (top panel) or cKO (bottom panel) cells at MD differentiation day 11.
(C) RT-qPCR for the myogenic genes *Myod1*, *Myog* and *Pax3* in parental and cKO cells transduced with retroviruses encoding Oct1, or empty vector controls. Transduced cells were selected using puromycin for 48 hr prior to differentiation for 11D. Empty vector values were set to 1 for both parental and cKO cells. Data represent an average of 3 independent biological replicates. Error bars depict ±SEM.
(D) Immunoblot showing ectopic Oct1 expression in parental and cKO cells. β-actin is shown as a loading standard.
(E) Sequential model of bivalency resolution for lineage-appropriate (MD-specific) genes. Pluripotent cells co-express Oct1 and Oct4, which co-bind to poised targets. Upon loss of

pluripotency and Oct4, Oct1 continues to occupy these genes. TGFβ signals allow for binding of Oct1 and Smad proteins to MD-specific targets, recruitment of Utx and demethylation of H3K27me3. Later, other transcription factors serve as primary "on" switches for muscle-specific gene expression.

## SUPPLEMENTAL FIGURE LEGENDS

**Figure S1. Expression of selected genes in UMAP clusters of pluripotent cells and cells at MD differentiation D3 and D6.**

Expression of eight different genes that highlight specific clusters is shown. Right: clusters are highlighted in undifferentiated, MD D3- and MD D6-differentiated cells similar to Figure 1A.

**Figure S2. Oct1 and Oct4 binding and co-binding in ESC cells (D0 of differentiation).**

(A) ChIP-seq signal tracks for *Polr2a* promoter are shown at left. Plot in center shows Oct1 and Oct4 ChIP-qPCR fold enrichment relative to a nonspecific 40S rRNA genomic region. At far right is Oct1→Oct4 sequential ChIP-qPCR (re-ChIP). qPCR data were normalized to a 40S rRNA nonspecific genomic region. An average of three biological replicates are shown. Error bars depict ±SEM. * *p*-value <0.05, ** *p*-value <0.01, *** *p*-value <0.001.

(B) Similar data for the *Pou5f1* enhancer.

(C) Similar data for the *Dll1* enhancer.

**Figure S3. Oct1 binding shifts from predominantly distal peaks in ESCs to promoter peaks in MD D6 differentiated cells.**

(A) Top enriched GO terms in MD D6 Oct1-bound peaks maintained during differentiation (shared with Oct1- and Oct4-bound peaks in ESCs, 807 peaks).

(B) Venn diagram similar to Figure 3A except showing common and unique binding peaks in MD D3-differentiated cells.

**Figure S4. Metabolic profiling of parental and cKO ESCs and MD-differentiated cells.**

(A) PCA plot of parental and cKO metabolic profiles.

(B) Heatmap of metabolites that change with MD differentiation of parental cells.

(C) Heatmap showing differential metabolites between parental and cKO cells at MD D6. 6 biological replicates were used for each condition.

## STAR METHODS

### RESOURCE AVAILABILITY

#### Data and Code Availability

The datasets generated during this study are available through the GEO website [GSE160941].

The code supporting the current study are available from the first author on request.

### METHOD DETAILS

#### Cell culture

ESCs were cultured as previously described (Shakya et al., 2015a) with 2i conditions: ERK inhibitor PD0325901 (1 μM, LC Laboratories) and GSK3 inhibitor CHIR99021 (3 μM, LC Laboratories). Cultures were maintained on irradiated feeders (ThermoFisher). Prior to all experiments ESCs were plated on gelatin to deplete the feeders. For MD differentiation, ESCs were plated on gelatin and cultured as previously described (Chal et al., 2015). Briefly, parental and cKO cells were cultured in N2B27 medium supplemented with recombinant Bmp4 (Peprotech) for 2 d. After 48 hr, media was changed to RDL (Rspo3, DMSO, LDN) medium. Cells were harvested 24 hr (day 3) or 96 hr (day 6) later. For muscle differentiation, cells were switched to HIFL (Hgf, Igf, Fgf, Ldn) medium and cultured for 48 hr (day 8) after which medium was switched to 2% horse serum (ThermoFisher). Cells were harvested on day 11 (overexpression experiments) or 19 (RT-qPCR).

#### Single cell RNA-seq

Single cell transcriptomes were analyzed as described previously (Dell'Orso et al., 2019). The 10X Genomics Chromium Single Cell Gene Expression Solution with 3' chemistry, version 3 (PN-1000075) was used to tag individual cells with 16 bp barcodes and specific transcripts with 10 bp Unique Molecular Identifiers (UMIs) according to manufacturer instructions. Briefly, single-cell suspensions were isolated using trypsinization and resuspension in PBS with 0.04% BSA (ThermoFisher). Suspensions were filtered through 40 μm cell strainers. Viability and cell count were assessed using a Countess I (ThermoFisher). Equilibrium to targeted cell recovery of 6,000 cells along with Gel Beads and reverse transcription reagents were loaded to Chromium Single Cell A to form Gel-bead-in Emulsions (GEMs). Within individual GEMs, cDNA generated from captured and barcoded mRNA was synthesized by reverse transcription at 53°C for 45 min. Samples were then heated to 85°C for 5 min. Subsequent A tailing, end repair, adaptor ligation and sample indexing were performed in bulk according to manufacturer instructions. The resulting

barcoding libraries were qualified on Agilent Technology 2200 TapeStation system and subjected to qPCR using a KAPA Biosystems Library Quantification Kit for Illumina Platforms (KK4842). The multiple libraries were then normalized and sequenced on an Illumina NovaSeq 6000 using the 2×150 PE mode.

**Data Processing and clustering**

Sequences from the Chromium platform were de-multiplexed and aligned using CellRanger ver. 3.1.0 (10X Genomics) with default parameters mm10-3.0.0. Clustering, filtering, variable gene selection and dimensionality reduction were performed using Seurat ver.3.1.5 (Stuart et al., 2019) according to the following workflow: 1, Cells with <300 detected genes and >10000 genes were excluded further analysis. 2, Cells with <12% UMIs mapping to mitochondrial genes were retained for downstream analysis. 3, The UMI counts per ten thousand were log-normalized for each cell using the natural logarithm. 4, Variable genes (2000 features) were selected using the FindVariableFeatures function. 5, Common anchors between the three parental timepoints (Fig.1A) or parental and cKO D6 (Fig.1C) were identified using FindIntegrationAnchors function that were further used to integrate these sets. 6, Gene expression levels in the integrated set were scaled along each gene and linear dimensional reduction was performed. The number of principal components was decided through the assessment of statistical plots (JackStrawPlot and ElbowPlot). 7, Cells were clustered using a by a shared nearest neighbor (SNN) modularity optimization-based clustering algorithm and visualized using two-dimensional uniform manifold approximation and projection (UMAP). 8, Cluster identities were defined based on the distribution of the specific markers. Differentiational gene expression analysis between the parental and cKO clusters was performed using FindMarkers. Genes with adjusted $p$<0.01 were marked red on scatter plots.

**Pseudotime and Velocity analysis**

Trajectory analysis of scRNA-seq was performed using Monocle v.2.16.0 (Trapnell et al., 2014). Parental and cKO sets were filtered using the same parameters as above and merged to generate WT and cKO sets. Cells were ordered based on gene lists for the ESC (beginning) and somite (end) clusters in parental UMAP (Fig.1A). Next, we performed dimensional reduction using the DDRTree method to visualize the dataset, ordered the cells by global gene expression levels, and visualized the trajectory of the dataset. Veolocity analysis was performed using the velocyto package (La Manno et al., 2018). Loom files were produced using following paramters: velocyto

run10x -m mm10. rmsk.gtf genes.gtf. Gtf files were produced from the Cell Ranger pipeline. Velocity embeddings were produced using the velocyto.r and SeuratWrappers packages. Matrices were filtered using following parameters: nFeature_spliced > 300, nFeature_spliced < 10000, nFeature_unspliced > 200, nFeature_unspliced < 6000, percent.mt < 12. Velocity was calculated using RunVelocity using following paremeters: deltaT = 1, kCells = 25, fit.quantile = 0.02. Velocity embedding were projected on T-SNE maps using the show.velocity.on.embedding.cor function.

**RT-qPCR**

RNA was isolated using TRIzol (ThermoFisher). cDNA was synthesized using a SuperScript Vilo cDNA Synthesis Kit (ThermoFisher). RT-qPCR oligonucleotide primers are listed in Table S6 and were confirmed to generate a single product of the correct size. To ensure specific PCR amplification, every RT-qPCR run was followed by a dissociation phase analysis (denaturation curve) to confirm the presence of single amplified peaks.

**ChIP**

ChIP-qPCR and ChIP-seq were performed as previously described (Perovanovic et al., 2016). Briefly, WT and cKO cells were crosslinked with 1% formaldehyde for 10 min and quenched for 5 min using 2.5M glycine. Culture plates were washed using ice cold PBS and cells were harvested by cell scaping. Cells were lysed in Farnham buffer (5 mM Pipes pH 8.0, 85 mM KCl, 0.5% NP-40) and subsequently in RIPA buffer (phosphate-buffered saline, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS). Chromatin was sheared using a Covaris sonicator for 5 min (30 sec on/30 sec off) with amplitude=40. Correct chromatin fragmentation was confirmed using 1% agarose gels. 50 µg of chromatin was subjected to IP overnight at 4°C with 4 µg of anti-Oct1 (Novus Biological), Oct4 (Santa Cruz) or H3K27me3 (Milipore) antibodies. As a control, we used 5 µg of sheared, non-precipitated input chromatin. Samples were incubated with protein G magnetic beads (ThermoFisher) for 5 hr and washed in Low Salt buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), High Salt buffer (identical but with 500 mM NaCl), LiCl buffer, and Tris-EDTA pH 8.0 plus 1 mM EDTA (TE buffer). Washes were performed at 4°C for 10 min with rotation. For re-ChIP, 2% fragmented chromatin was saved as input and the rest used for IP with Oct4 antibody at 4°C overnight on a rotator. Samples were then incubated with magnetic beads for 5h at 4°C. The beads were washed for 10 min with Low Salt buffer, High Salt buffer, Low Salt buffer, LiCl buffer, and TE buffer sequentially at 4°C. Chromatin

was eluted with 300 $\mu$L IP Elution buffer at RT, then diluted 10-fold in RIPA buffer. Diluted samples were then subjected to a second IP with 4 $\mu$g of Oct1 antibody (Novus Biological) at 4°C overnight, and incubated with magnetic beads for 5 hr at 4°C. The beads were washed again as described above, then eluted with 300 $\mu$L IP Elution Buffer at RT. Re-ChIP samples, together with the 2% input, were incubated at 65°C overnight to reverse crosslinking. DNA was purified using phenol-chloroform-isoamyl alcohol extraction followed by PCR clean up. qPCR primers can be found in Table S6 and were confirmed to generate a single product of the correct size. The results were reported as qPCR values normalized to input chromatin (gDNA) and non-specific region and presented as fold enrichment.

**ChIP-seq analysis**

After chromatin was precipitated as described above, and libraries were sequenced using Illumina NovaSeq. Between 22 and 26 million paired-end Illumina sequencing reads were aligned to the mouse *Mm10* reference genome using Novocraft novoalign v3.8.2, allowing for one random alignment of multi-mapping reads, and providing the adapter sequences for automatic trimming concordant with alignment. ChIP was analyzed using the MultiRepMacsChIPSeq pipeline v12.2, using options "--pe --optdist 10000 --dupfrac 0.15 --mapq 10 --cutoff 2 --tdep 10 --peaksize 500 --peakgap 100 --species mouse --chrskip 'chrMIPhiX' --blacklist mm10.blacklist.bed".

**Immunoprecipitation**

Cells were lysed with Cell Lysis Buffer (Life Technologies) in the presence of protease inhibitors (EDTA-free tablet, Roche). IP was performed using 500 µg of extract. Extracts were incubated with 4 µg of anti-Utx (Cell Signaling, D3Q1I) or Oct1 (Novus Biologicals, NBP2-21584) antibodies, or rabbit IgG control overnight at 4°C. Protein-antibody complexes were precipitated with protein-G magnetic beads (Thremo Fisher) for 3 hr at 4°C with rotation and washed 3 times with Low Salt buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100) plus protease inhibitors. Precipitated proteins were analyzed by immunoblot.

**Bulk RNA-seq**

RNA was prepared from three independent cultures of undifferentiated or 3 d and 6 d MD-differentiated parental or cKO ESCs. 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 and RNA UD Indexes. Molarity

of adapter-modified molecules was defined by qPCR using the Kapa Biosystems Library Quant Kit. Individual libraries were normalized to 1.3 nM. Sequencing libraries were chemically denatured and applied to an Illumina NovaSeq flow cell using the NovaSeq XP chemistry workflow. Following transfer of the flowcell to an Illumina NovaSeq instrument, 2×51 cycle paired-end sequence was performed using a NovaSeq S1 reagent kit. Between 13 and 18 million paired-end reads were generated for each condition. More than 99% of aligned reads mapping to the correct strand.

**Bulk RNA-seq analysis**

The Illumina adapter sequence was trimmed using cutadapt version 1.16. Fastq data quality were checked using Fastqc verision 0.11.5. Reads were aligned to the mouse *Mm10* genome using STAR version 2.7.3a in two-pass mode. Aligned reads were checked for quality using the Picard tools' CollectRnaSeqMetrics command to count the number of read-matching exons, UTRs, introns and intergenic regions, and to calculate normalized gene coverage from the top 1000 expressed transcripts. Between 13 and 18 million paired-end reads were generated for each condition, with >99% of aligned reads mapping to the correct strand. Differentially expressed genes were identified using a 5% FDR with DESeq2 version 1.24.0 (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 $p<0.01$ were selected as differentially expressed. Figures were generated in R version 4.0.0 using functions from ggplots libraries and pheatmap.

**Luciferase reporter assay**

Single or combination mutations were introduced into the Oct1 and Smad consensus binding sites in the following mouse *Myog* regulatory element (*Mm10* chr1:134,285,666-134,285,750). IDT g-blocks® were synthesized to contain WT sequences or single or combined mutations in the Oct1 or Smad binding sites, fused upstream of the CMV basal promoter (-216-13 relative to TSS). G-blocks were inserted using sequence- and ligase-independent cloning (Li and Elledge, 2012) upstream of the coding sequence for a secreted nano-luciferase following digestion of vector pNL2.3 (Promega) using *Eco*RV and *Hind*III. The veracity of the cloned inserts was confirmed by Sanger sequencing. 200 ng of reporter plasmid were co-transfected into Oct1-deficient MEFs (Shakya et al., 2009) in DMEM media lacking phenol red (ThermoFisher) together with 400 ng MMP9-mCherry (Wider and Picard, 2017) in xxx ng total transfected DNA. Where indicated, 400 ng pBabePuro-Oct1 was included in the transfection mix. pUC18 plasmid comprised the balance

of the transfected DNA. Where indicated, transfected cells were provided with 5 ng recombinant mouse TGFb1 protein (R&D Systems). mCherry fluorescence was determined first by exciting at 570 nm and measuring emission at 610 nm with a 100 msec time delay using an Envision Xcite Multilabel Plate Reader . Luminescence was measured using Nano-Glo Luciferase (Promega) and a Modulus luminescence plate reader.

**Immunofluorescence**

Immunofluorescence was performed as described previously (Gnocchi et al., 2009) with modifications, using rabbit anti-H3K27me3 (Milipore) and mouse anti-MyHC-emb (eMyHC, Developmental Hybridoma bank) antibodies. Secondary antibodies were goat anti-rabbit-Alexa568 and anti-mouse-Alexa568 (ThermoFisher).

**Retroviral Oct1 overexpression**

Oct1 was ectopically expressed in ESC cells using a previously described retroviral vector (pBabePuro-Oct1) (Kang et al., 2009b). pBabePuro was used as an empty vector control. The vector was co-transfected together with the pCL-Eco packaging plasmid into HEK293T cells to generate retroviral particles. Retroviral supernatants were collected 48 hr later, filtered through 0.45 μm filters and applied to ESCs cultures maintained on puromycin-resistant feeder fibroblasts (ThermoFisher). The mixed population of cells was subjected to selection with puromycin for 48 hr.

**Metabolic profiling**

Cold 90% methanol (MeOH) was added to each sample to give a final concentration of 80%. Samples were then incubated at -20˚C for 1 hr. After incubation the samples were centrifuged at 20,000×*g* for 10 min at 4˚C. The supernatant was then transferred from each sample tube into a labeled, fresh micro centrifuge tube. Pooled quality control samples were made by removing a fraction of collected supernatant from each sample and process blanks were made using only extraction solvent and no cell culture. The samples were then dried en vacuo. GC-MS was performed with an Agilent 5977b GC-MS MSD-HES and an Agilent 7693A automatic liquid sampler. Data were analyzed using in-house software to prepare for analysis by the "MetaboAnalyst" software tool (Chong et al., 2018). Statistical analysis was performed using MetaboAnalystR.

## ACKNOWLEDGEMENTS

We thank G. Kardon and C. Kikani for critical reading of the manuscript. We thank B. Dalley and O. Allen at the HCI High-Throughput Genomics facility and T. Parnell and B. Lohman at the HCI Bioinformatic Analysis Shared Resource for assistance with ChIP-seq and scRNA-seq. We thank James E. Cox and Tyler Van Ry from the Metabolomics Core for their assistance with metabolomic profiling. We thank Olivier Pourquié from the Harvard Medical School for the assistance with the mesodermal differentiation protocols. We thank D. Picard and D. Wider for the MMP9-mCherry construct. This work was supported by a grant from the National Institutes of Health/National Institute of General Medical Sciences (R01GM122778) to DT.

*Author contributions:* DT conceived the study, and provided administrative and material support. JP conceived and supervised experiments, designed experiments and acquired and interpreted data. YW, ZS acquired and interpreted data. MBC generated reagents and analyzed data. All authors were involved in writing, reviewing and revising the manuscript.

804 **REFERENCES**

805

806 Abascal, F., Acosta, R., Addleman, N.J., Adrian, J., Afzal, V., Aken, B., Akiyama, J.A., Jammal,
807 O. Al, Amrhein, H., Anderson, S.M., et al. (2020). Expanded encyclopaedias of DNA elements in
808 the human and mouse genomes. Nature *583*, 699–710.
809 Bain, G., Kitchens, D., Yao, M., Huettner, J.E., and Gottlieb, D.I. (1995). Embryonic stem cells
810 express neuronal properties in vitro. Dev. Biol. *168*, 342–357.
811 Beddington, R.S.P., and Robertson, E.J. (1989). An assessment of the developmental potential
812 of embryonic stem cells in the midgestation mouse embryo. Development *105*, 733–737.
813 Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner,
814 A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental
815 genes in embryonic stem cells. Cell *125*, 315–326.
816 Chal, J., Oginuma, M., Tanoury, Z. Al, Gobert, B., Sumara, O., Hick, A., Bousson, F., Zidouni,
817 Y., Mursch, C., Moncuquet, P., et al. (2015). Differentiation of pluripotent stem cells to muscle
818 fiber to model Duchenne muscular dystrophy. Nat. Biotechnol. *33*, 962–969.
819 Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S., and Xia, J. (2018).
820 MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. Nucleic
821 Acids Res. *46*, W486–W494.
822 Cliff, T.S., Wu, T., Boward, B.R., Yin, A., Yin, H., Glushka, J.N., Prestegaard, J.H., and Dalton,
823 S. (2017). MYC Controls Human Pluripotent Stem Cell Fate Decisions through Regulation of
824 Metabolic Flux. Cell Stem Cell *21*, 502–516.e9.
825 Cloos, P.A.C., Christensen, J., Agger, K., and Helin, K. (2008). Erasing the methyl mark:
826 Histone demethylases at the center of cellular differentiation and disease. Genes Dev. *22*,
827 1115–1140.
828 Collinson, A., Collier, A.J., Morgan, N.P., Sienerth, A.R., Chandra, T., Andrews, S., and Rugg-
829 Gunn, P.J. (2016). Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-
830 Renewal and Differentiation Defects in Human Embryonic Stem Cells. Cell Rep.
831 Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson,
832 E.J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive
833 streak in the mouse. Trends Genet. *10*, 308–308.
834 Dell'Orso, S., Juan, A.H., Ko, K.-D., Naz, F., Perovanovic, J., Gutierrez-Cruz, G., Feng, X., and
835 Sartorelli, V. (2019). Correction: Single cell analysis of adult mouse skeletal muscle stem cells in
836 homeostatic and regenerative conditions (doi: 10.1242/dev.174177). Development *146*,

dev181743.

Dhar, S.S., Lee, S.H., Chen, K., Zhu, G., Oh, W.K., Allton, K., Gafni, O., Kim, Y.Z., Tomoiga, A.S., Barton, M.C., et al. (2016). An essential role for UTX in resolution and activation of bivalent promoters. Nucleic Acids Res. *44*, 3659–3674.

Diaz-Cuadros, M., Wagner, D.E., Budjan, C., Hubaud, A., Tarazona, O.A., Donelly, S., Michaut, A., Al Tanoury, Z., Yoshioka-Kobayashi, K., Niino, Y., et al. (2020). In vitro characterization of the human segmentation clock. Nature *580*, 113–118.

Gnocchi, V.F., White, R.B., Ono, Y., Ellis, J.A., and Zammit, P.S. (2009). Further characterisation of the molecular signature of quiescent and activated mouse muscle satellite cells. PLoS One *4*, e5205.

Guibentif, C., Griffiths, J.A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V., Göttgens, B., and Marioni, J.C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo. Dev. Cell *56*, 141–153.e6.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Kang, J., Gemberling, M., Nakamura, M., Whitby, F.G., Handa, H., Fairbrother, W.G., and Tantin, D. (2009a). A general mechanism for transcription regulation by Octl and Oct4 in response to genotoxic and oxidative stress. Genes Dev. *23*, 208–222.

Kang, J., Gemberling, M., Nakamura, M., Whitby, F.G., Handa, H., Fairbrother, W.G., and Tantin, D. (2009b). A general mechanism for transcription regulation by Oct1 and Oct4 in response to genotoxic and oxidative stress. Genes Dev. *23*, 208–222.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genet. *4*, e1000242.

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: Transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438–2444.

Li, M.Z., and Elledge, S.J. (2012). SLIC: A method for sequence- and ligation-independent

cloning. Methods Mol. Biol. *852*, 51–59.

Liu, Z., Lin, X., Cai, Z., Zhang, Z., Han, C., Jia, S., Meng, A., and Wang, Q. (2011). Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. J. Biol. Chem. *286*, 28520–28532.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

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

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

Oginuma, M., Moncuquet, P., Xiong, F., Karoly, E., Chal, J., Guevorkian, K., and Pourquié, O. (2017). A Gradient of Glycolytic Activity Coordinates FGF and Wnt Signaling during Elongation of the Body Axis in Amniote Embryos. Dev. Cell *40*, 342–353.e10.

Palmieri, S.L., Peter, W., Hess, H., and Schöler, H.R. (1994). Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the firsttwo extraembryonic cell lineages involved in implantation. Dev. Biol. *166*, 259–267.

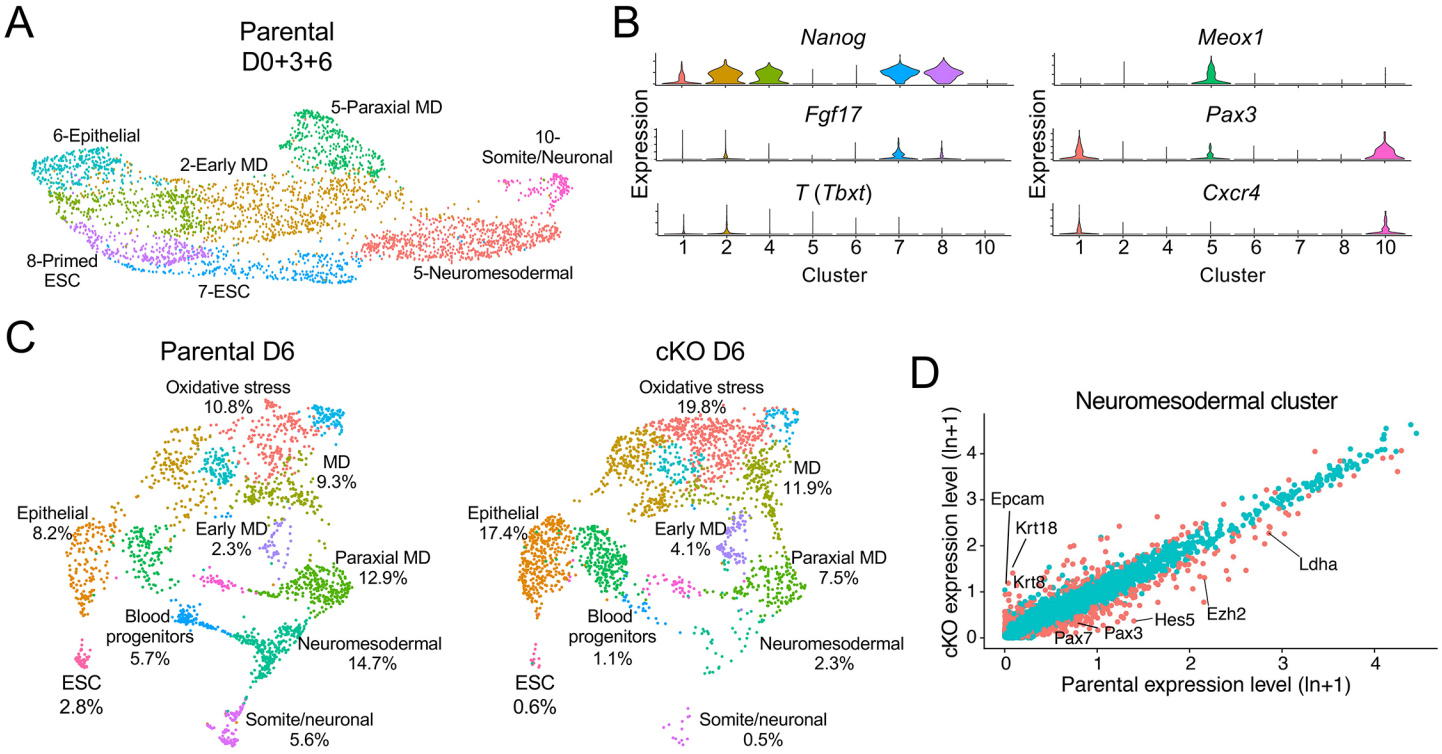
Perovanovic, J., DellOrso, S., Gnochi, V.F., Jaiswal, J.K., Sartorelli, V., Vigouroux, C., Mamchaoui, K., Mouly, V., Bonne, G., and Hoffman, E.P. (2016). Laminopathies disrupt epigenomic developmental programs and cell fate. Sci. Transl. Med. *8*, 335ra58-335ra58.

Schlesinger, S., and Meshorer, E. (2019). Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. Dev. Cell *48*, 135–150.

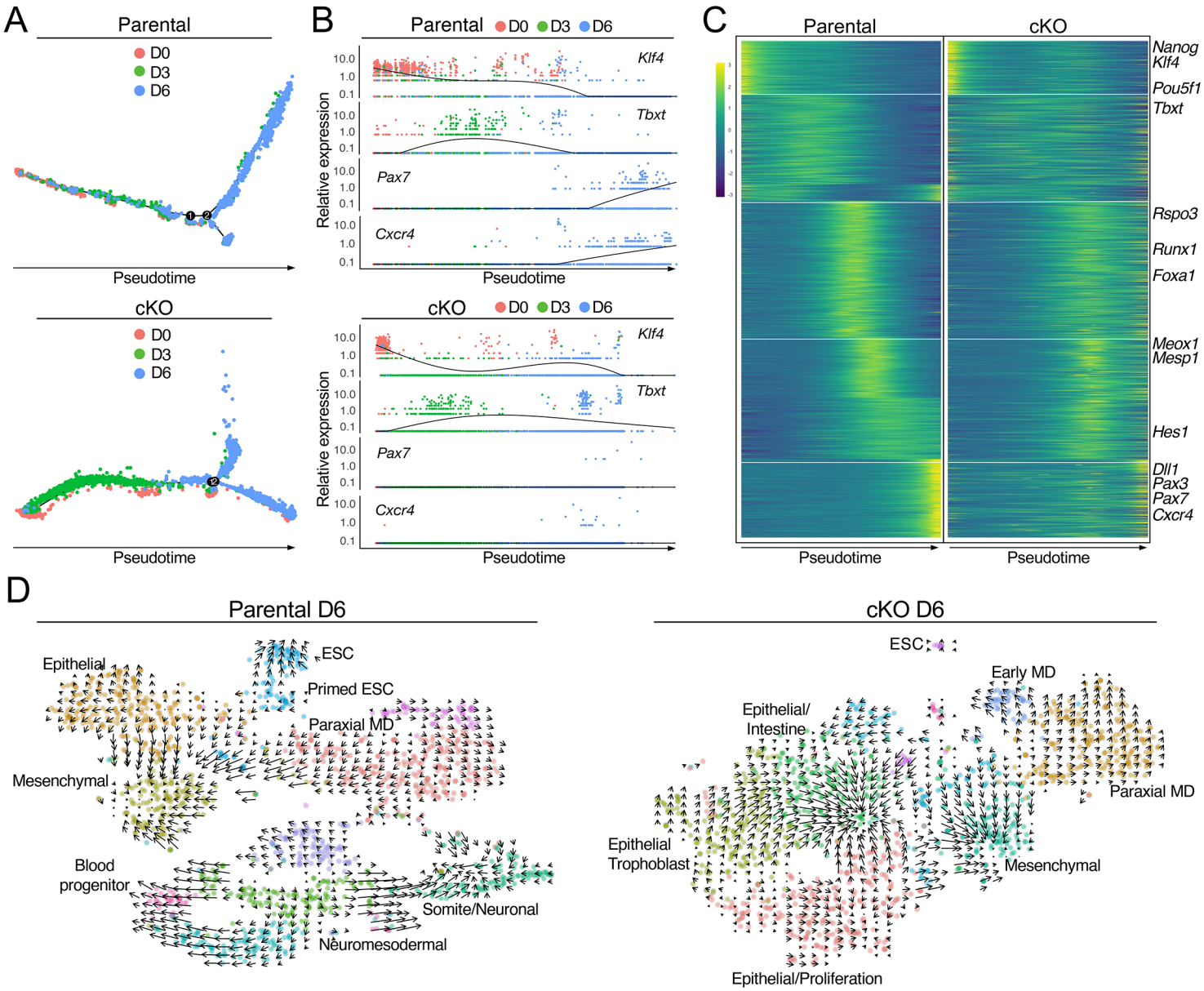
Sebastiano, V., Dalvai, M., Gentile, L., Schubart, K., Sutter, J., Wu, G.M., Tapia, N., Esch, D., Ju, J.Y., Hübner, K., et al. (2010). Oct1 regulates trophoblast development during early mouse

903 embryogenesis. Development *137*, 3551–3560.
904 Seydoux, G., and Braun, R.E. (2006). Pathway to Totipotency: Lessons from Germ Cells. Cell
905 *127*, 891–904.
906 Shakya, A., Cooksey, R., Cox, J.E., Wang, V., McClain, D.A., and Tantin, D. (2009). Oct1 loss
907 of function induces a coordinate metabolic shift that opposes tumorigenicity. Nat. Cell Biol. *11*,
908 320–327.
909 Shakya, A., Kang, J., Chumley, J., Williams, M.A., and Tantin, D. (2011). Oct1 is a switchable,
910 bipotential stabilizer of repressed and inducible transcriptional states. J. Biol. Chem. *286*, 450–
911 459.
912 Shen, Z., Kang, J., Shakya, A., Tabaka, M., Jarboe, E.A., Regev, A., and Tantin, D. (2017).
913 Enforcement of developmental lineage specificity by transcription factor Oct1. Elife *6*.
914 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y.,
915 Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell
916 Data. Cell *177*, 1888–1902.e21.
917 Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse
918 Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.
919 Tantin, D. (2013). Oct transcription factors in development and stem cells: insights and
920 mechanisms. Development *140*, 2857–2866.
921 Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak,
922 K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions
923 are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.
924 Wang, V.E.H., Schmidt, T., Chen, J., Sharp, P.A., and Tantin, D. (2004). Embryonic Lethality,
925 Decreased Erythropoiesis, and Defective Octamer-Dependent Promoter Activation in Oct-1-
926 Deficient Mice. Mol. Cell. Biol. *24*, 1022–1032.
927 Wider, D., and Picard, D. (2017). Secreted dual reporter assay with Gaussia luciferase and the
928 red fluorescent protein mCherry. PLoS One *12*, e0189403.
929 Zhang, Y., Lahmann, I., Baum, K., Shimojo, H., Mourikis, P., Wolf, J., Kageyama, R., and
930 Birchmeier, C. (2021). Oscillations of Delta-like1 regulate the balance between differentiation
931 and maintenance of muscle stem cells. Nat. Commun. *12*.
932 Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L.M., and Kuehn, M.R. (1993). Nodal is a novel TGF-
933 β-like gene expressed in the mouse node during gastrulation. Nature *361*, 543–547.
934

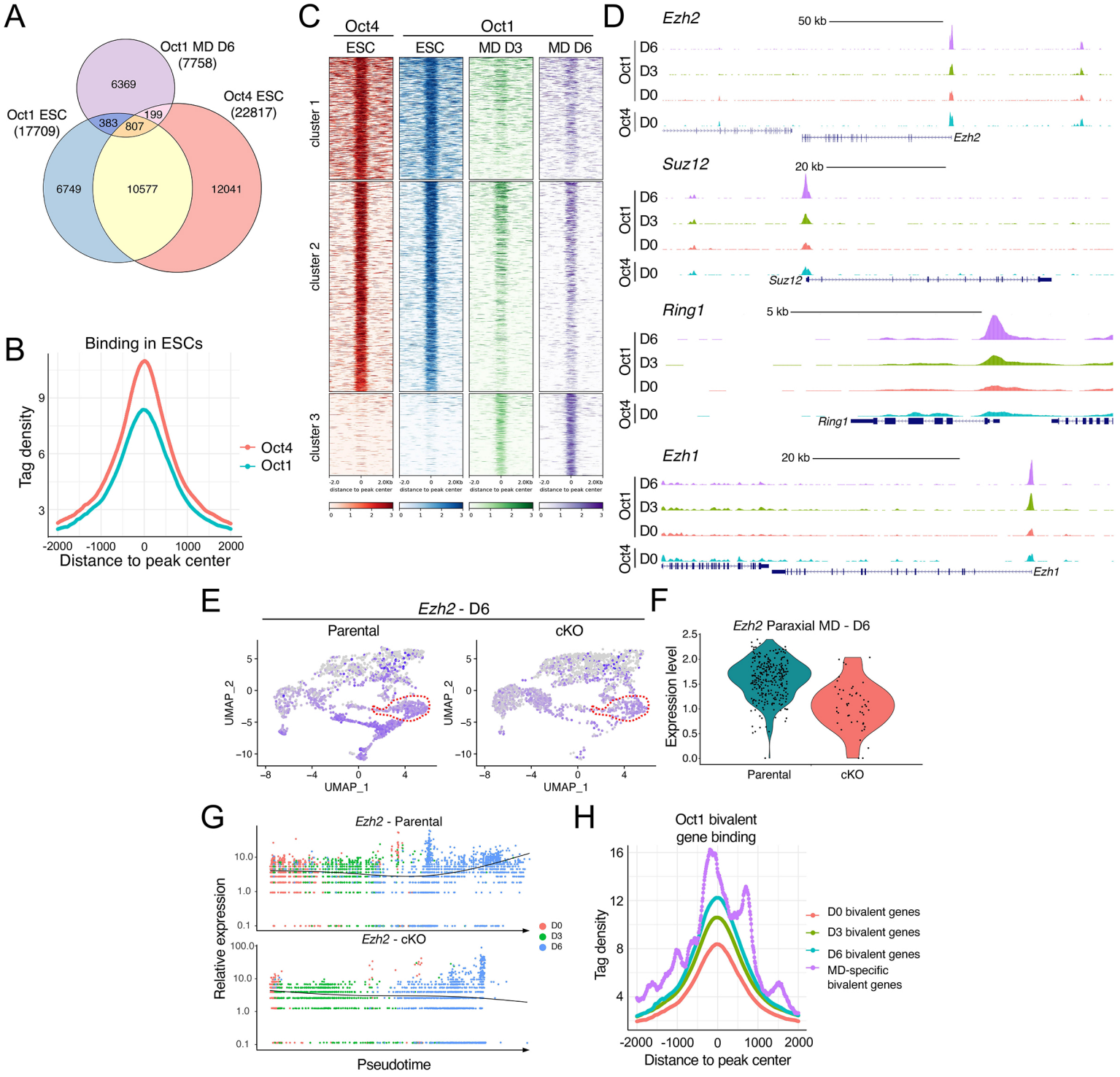
# Perovanovic *et al.* Figure 1

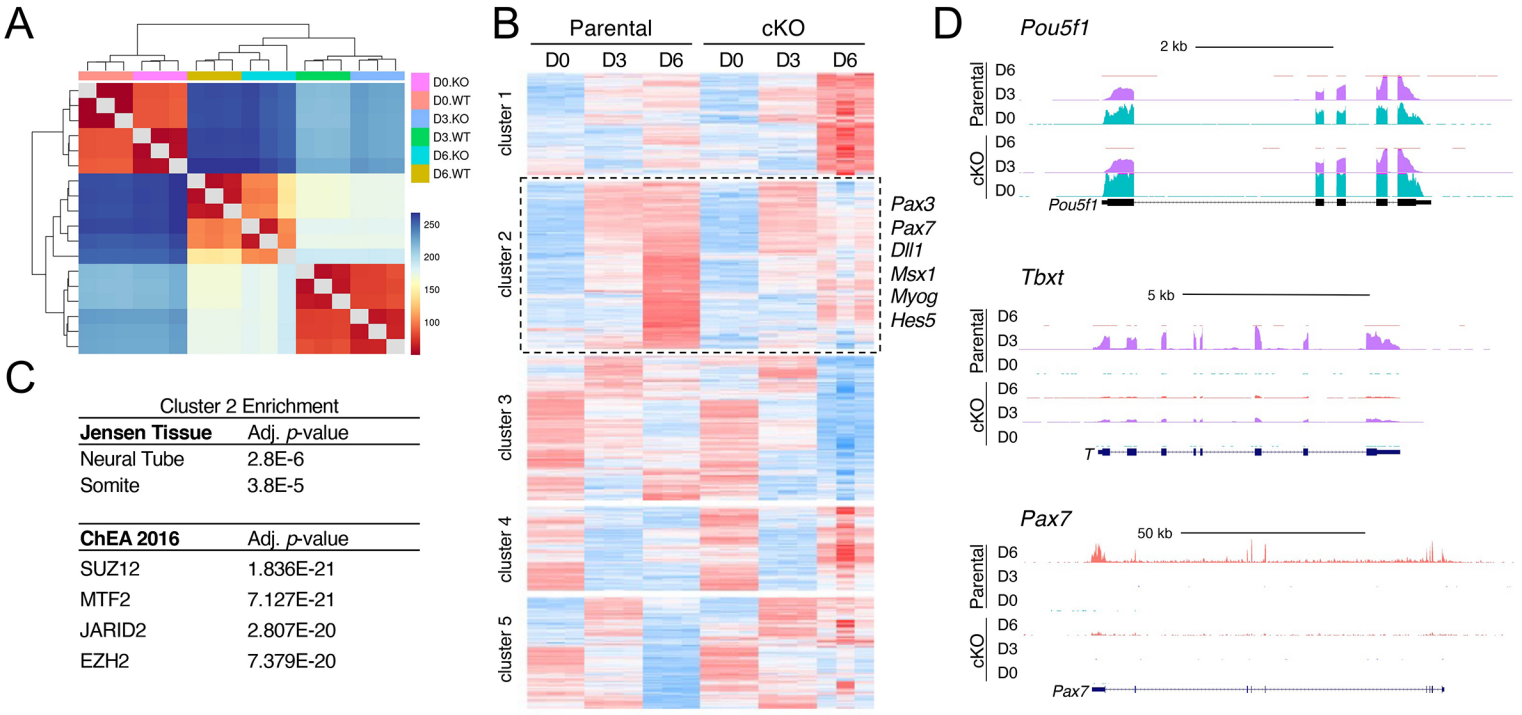


# Perovanovic *et al.* Figure 2



# Perovanovic *et al.* Figure 3



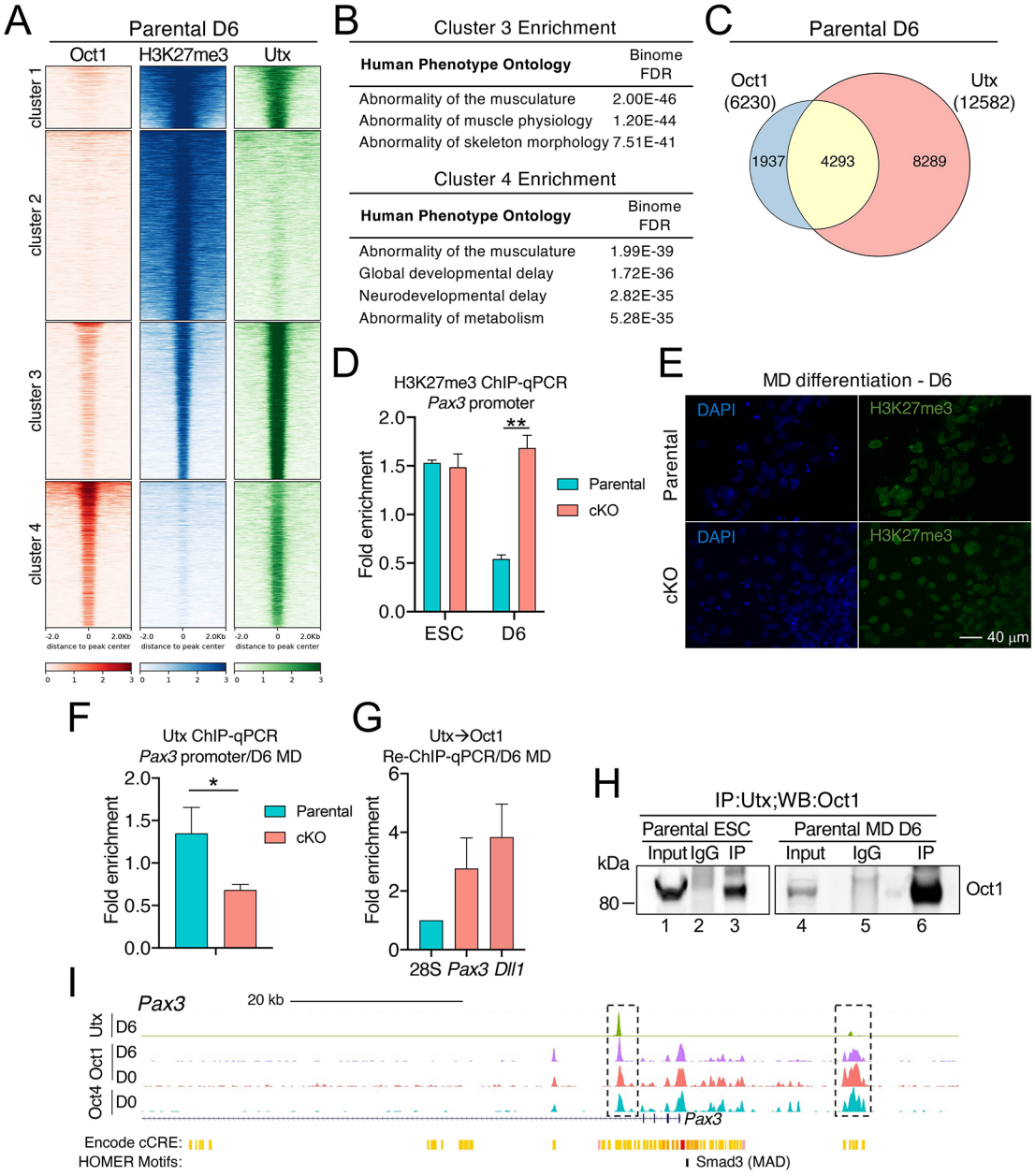


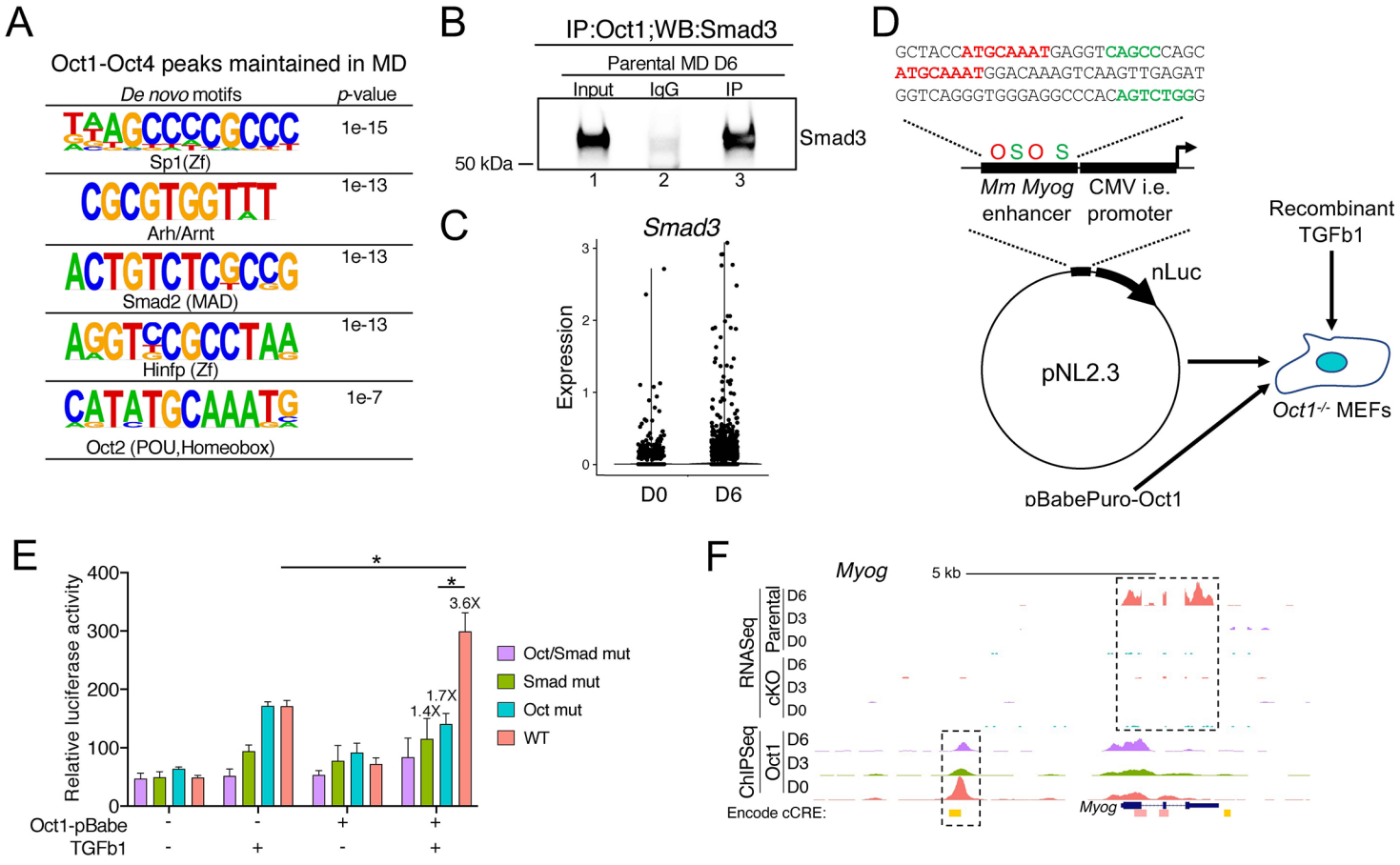

C

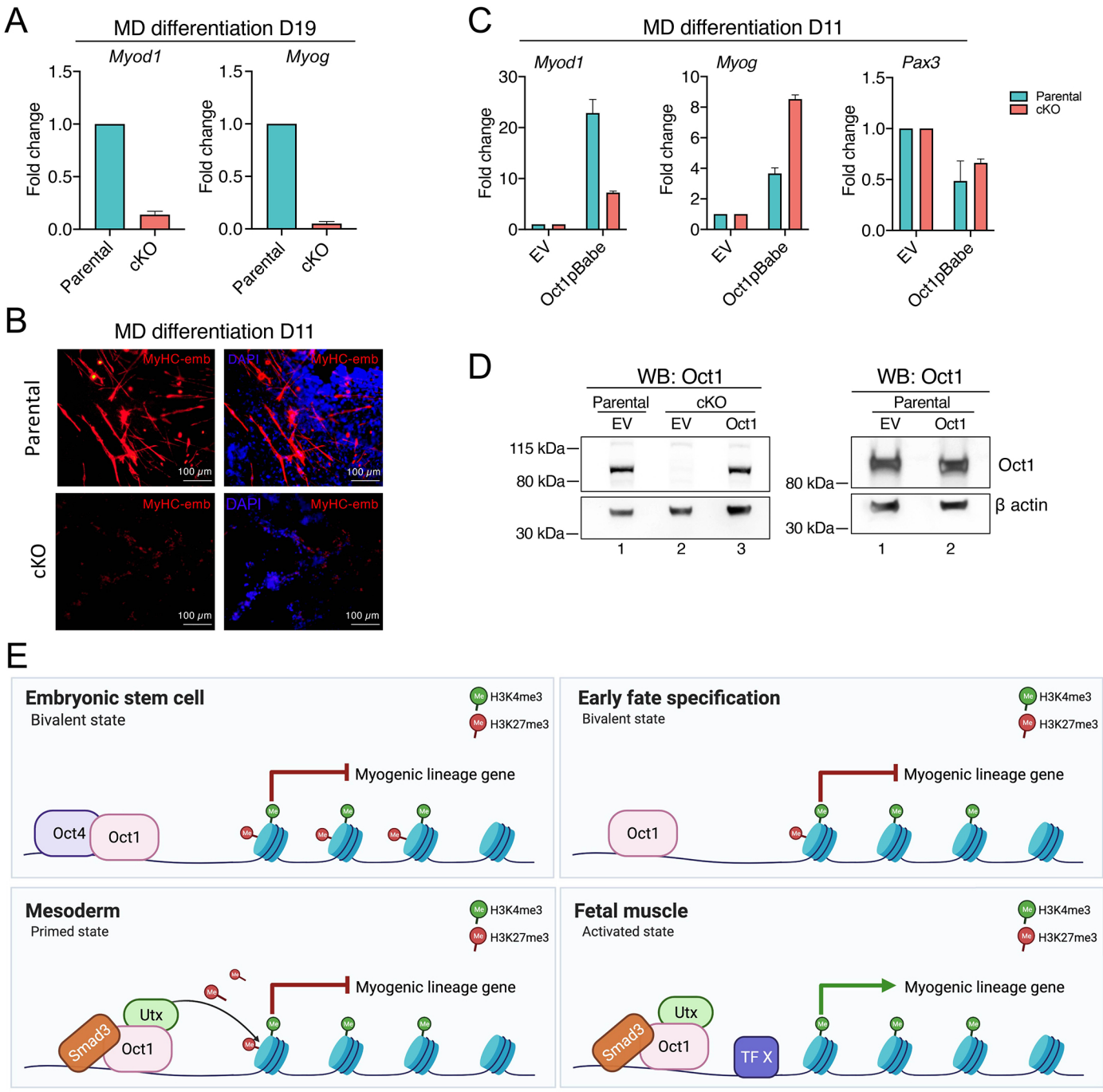
Cluster 2 Enrichment

| **Jensen Tissue** | Adj. *p*-value |
|---|---|
| Neural Tube | 2.8E-6 |
| Somite | 3.8E-5 |

| **ChEA 2016** | Adj. *p*-value |
|---|---|
| SUZ12 | 1.836E-21 |
| MTF2 | 7.127E-21 |
| JARID2 | 2.807E-20 |
| EZH2 | 7.379E-20 |







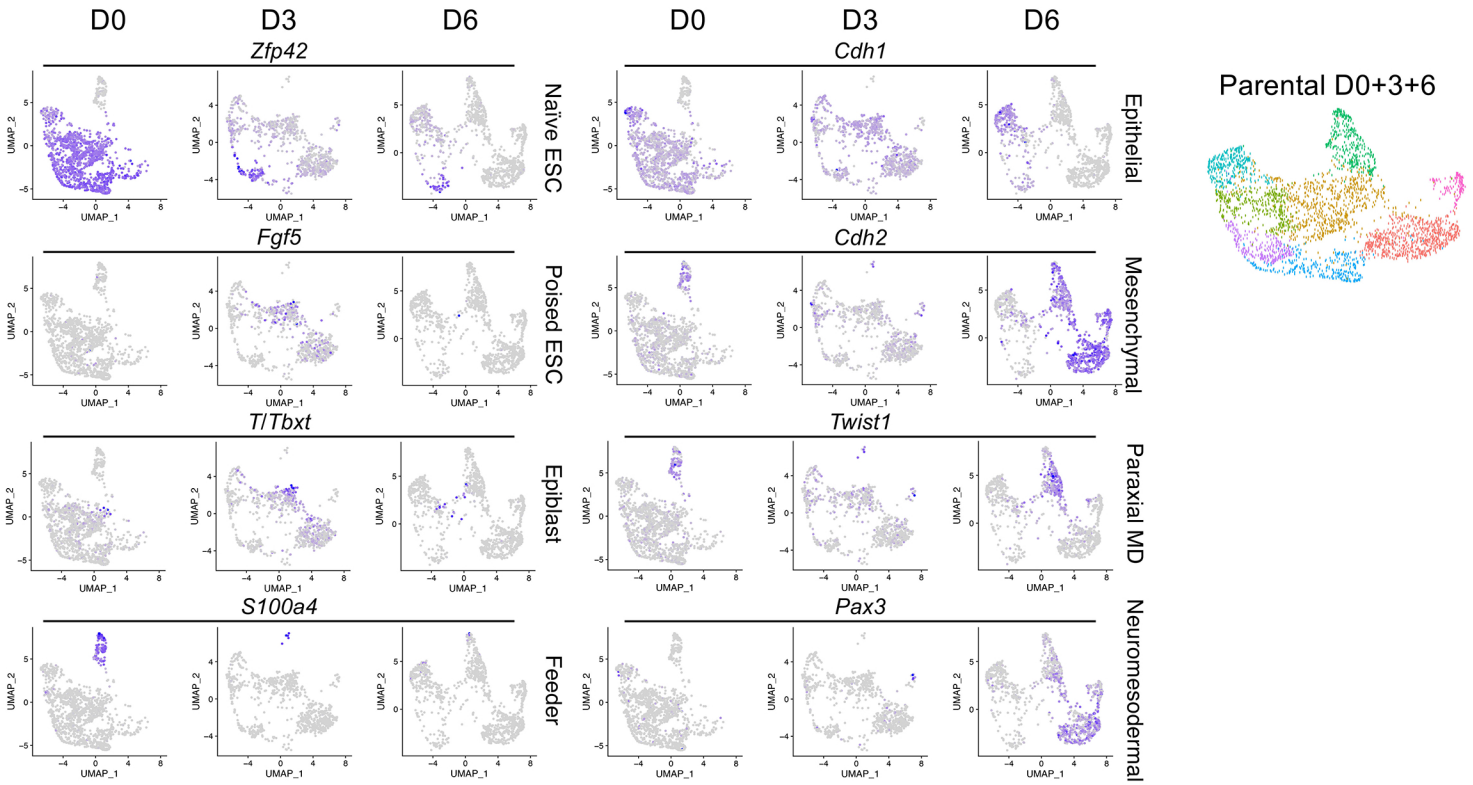



**Figure S1. Expression of selected genes in UMAP clusters of pluripotent cells and cells at MD differentiation D3 and D6.**

Expression of eight different genes that highlight specific clusters is shown. Right: clusters are highlighted in undifferentiated, MD D3- and MD D6-differentiated cells similar to Figure 1A.

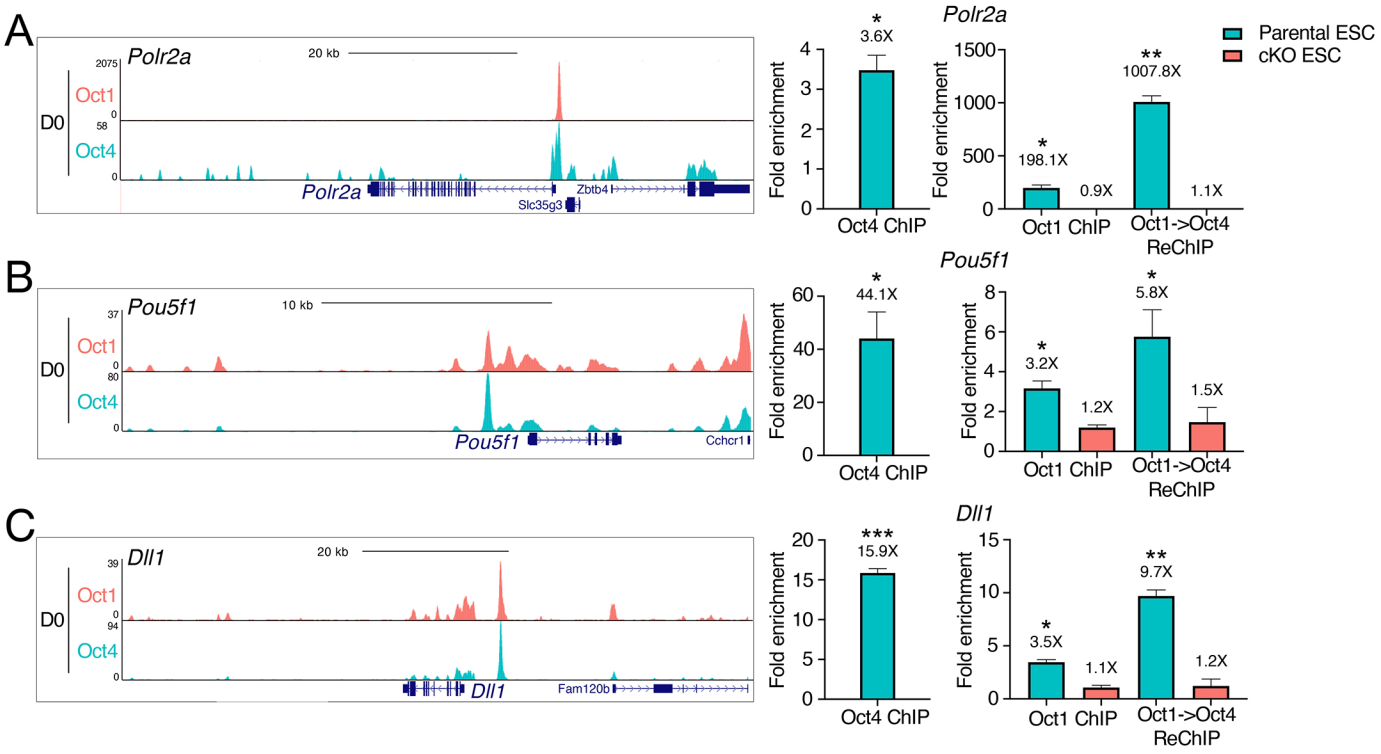



**Figure S2. Oct1 and Oct4 binding and co-binding in ESC cells (D0 of differentiation).**
(A) ChIP-seq signal tracks for *Polr2a* promoter are shown at left. Plot in center shows Oct1 and Oct4 ChIP-qPCR fold enrichment relative to a nonspecific 40S rRNA genomic region. At far right is Oct1→Oct4 sequential ChIP-qPCR (re-ChIP). qPCR data were normalized to a 40S rRNA nonspecific genomic region. An average of three biological replicates are shown. Error bars depict ±SEM. * *p*-value <0.05, ** *p*-value <0.01, *** *p*-value <0.001.
(B) Similar data for the *Pou5f1* enhancer.
(C) Similar data for the *Dll1* enhancer.

A

D6 Mesoderm

| GO Biological Process (807 peaks) | Binom FDR *p*-Adj |
|---|---|
| response to oxidative stress | 0.001836177 |
| ribosomal small subunit biogenesis | 0.001804605 |
| mitochondrial membrane organization | 0.014115737 |
| response to reactive oxygen species | 0.020594866 |
| response to hydrogen peroxide | 0.024908306 |
| cellular response to arsenic-containing substance | 0.034804788 |
| regulation of Notch signaling pathway | 0.0362016 |

| GO Cellular Component (807 peaks) | Binom FDR *p*-Adj |
|---|---|
| PRC1 complex | 3.338E-06 |
| cleavage furrow | 0.00029203 |
| cell division site part | 0.0003536 |
| PcG protein complex | 0.00040775 |
| collagen trimer | 0.00768286 |
| cluster of actin-based cell projections | 0.03337456 |
| Golgi stack | 0.04605178 |

| MGI Phenotype (807 peaks) | Binom FDR *p*-Adj |
|---|---|
| embryonic lethality between implantation and somite formation | 8.2454E-05 |
| embryonic lethality between implantation and placentation | 8.9031E-05 |

B

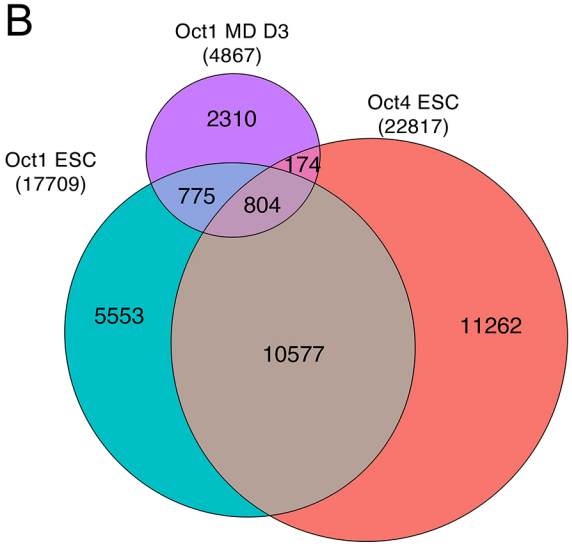



**Figure S3. Oct1 binding shifts from predominantly distal peaks in ESCs to promoter peaks in MD D6 differentiated cells.**

(A) Top enriched GO terms in MD D6 Oct1-bound peaks maintained during differentiation (shared with Oct1- and Oct4-bound peaks in ESCs, 807 peaks).

(B) Venn diagram similar to Figure 3A except showing common and unique binding peaks in MD D3-differentiated cells.

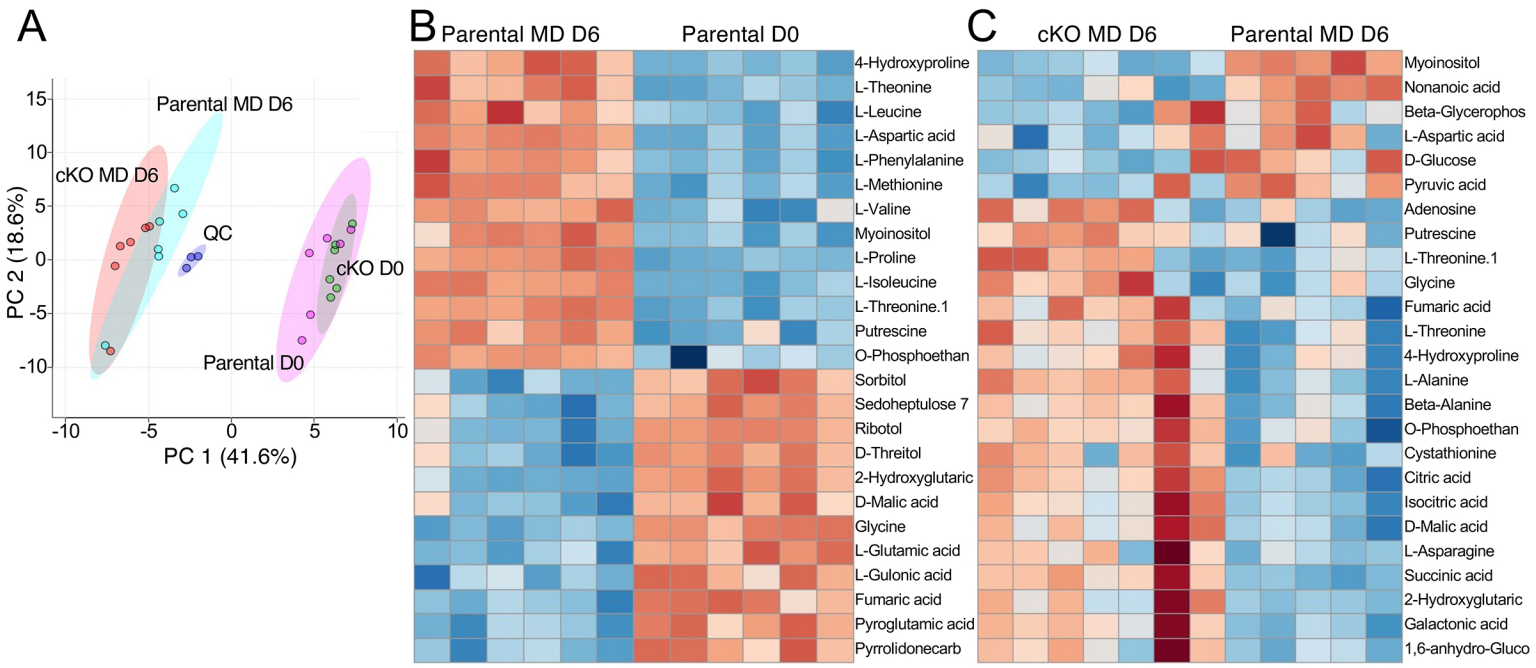



**Figure S4. Metabolic profiling of parental and cKO ESCs and MD-differentiated cells.**
(A) PCA plot of parental and cKO metabolic profiles.
(B) Heatmap of metabolites that change with MD differentiation of parental cells.
(C) Heatmap showing differential metabolites between parental and cKO cells at MD D6. 6 biological replicates were used for each condition.