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.

*Corresponding author
email: dean.tantin@path.utah.edu

Short title: Oct1 mediates lineage-specific gene activation

Keywords pre-implantation development; embryonic stem cells; mesoderm; single-cell RNA sequencing; Oct1/Pou2f1; Utx/Kmd3a; Smad3

Characters in title (including spaces): 82
Words in abstract: 157
Words main text: 4,562
Character count (including spaces): 56,952
Character count (excluding spaces): 48,701
Total main figures and tables: 7

Abstract

The pathways used by cells to transition between undifferentiated, pluripotent state and tissue-specific states are incompletely understood. Here we show that the widely-expressed transcription factor Oct1/Pou2f1 activates silent, developmental lineage-appropriate genes to "canalize" developmental progression. Using Oct1 inducible knockout embryonic stem cells, we show that that Oct1 deficiency impairs mesodermal and terminal muscle differentiation in a manner that can be rescued by Oct1 retroviral expression. Using bulk RNA-seq, we show that mesoderm-specific genes are not correctly induced early in the differentiation timecourse. Single-cell gene expression profiling reveals that Oct1-deficient cells lose coherence in temporal induction of lineage programs, and show inappropriate developmental lineage branching resulting in poorly differentiated cells state with epithelial characteristics and hallmarks of oxidative stress. In embryonic stem cells, Oct1 co-binds with Oct4 to genes critical for mesoderm induction. The Utx/Kdm6a histone lysine demethylase also binds to many of these genes, and using a prototypic *Pax3* gene we show that Oct1 recruits Utx to remove inhibitory H3K27me3 marks and activate expression. The specificity of the ubiquitous Oct1 protein for mesodermal genes can be explained by cooperative interactions with lineage-driving Smad transcription factors, as we show that Smad and Oct binding sites frequently coexist mesoderm-specific genes, that Oct1 and Smad3 interact, and that the sites and factors act cooperatively at the *Myog* enhancer. Overall, these results identify Oct1 as a key mediator of the induction of mesoderm lineage-specific genes.

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 embryonic 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 the multilayered post-implantation embryo. 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 lack DNA methylation and are poised for activation. During development, gene bivalency is thought to resolve via either removal of activating marks and gene silencing, or removal of repressive marks and gene activation. Which poised 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 activation of the correct genes 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 silenced in their differentiated progeny before bivalency resolution and the induction of tissue- and cell type-specific gene expression (DeVeale et al., 2013). 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; V. E. H. 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 proliferate and self-renewal normally, but upon differentiation show phenotypic and gene expression defects including decreased developmental lineage-specific gene expression and elevated expression of developmentally incorrect genes (Shen et al., 2017). Extended incubation of these cells supports an interpretation that there is a failure to differentiate rather than a simple kinetic delay (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 pluripotent cells lacking Oct1 fail to form terminally differentiated myotubes. Differentiation is restored by Oct1 retroviral complementation. Bulk RNA-seq profiling early in the differentiation timecourse identifies gene expression abnormalities early in MD differentiation that predict the later phenotypic defects. Single-cell RNA sequencing reveals that cells lacking Oct1 accumulate abnormal cell populations associated with epithelial characteristics, oxidative stress and early MD, while almost completely failing to reach somite-stage differentiation. Pseudotime analysis reveal increased predilection to proceed down incorrect developmental trajectories, and “fuzzy” differentiation programs. We show that Oct1 interacts with components of the Utx (Kdm3a) H3K27me3 demethylase complex, and recruits Utx to developmentally appropriate target genes. At tested targets this results in specific removal of H3K27me3 and de-repression. 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 activating potential to MD-specific genes. The cumulative results support a model in which Oct1 “canalizes” differentiation by supporting the induction of lineage-specific genes.

Results

Loss of Oct1 causes aberrant mesodermal differentiation

To study the consequences of Oct1 deficiency for MD development, we differentiated Oct1 inducible-conditional *Pou2f1*$^{fl/fl}$;Rosa26-CreER^{T2};LSL-YFP control ESCs (hereafter, "parental" cells) and derived tamoxifen-treated, Oct1-deficient cells (Shen et al., 2017) (hereafter, "cKO") using an MD differentiation protocol that results in differentiated myotubes after 10 days 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). Parental cells were differentiated for 11 days (D11) to generate early myotubes to identify changes in myotube structure. Immunostaining for embryonic myosin heavy chain (MyH-emb) reveals robust expression in fused myotubes in parental cells, with cKO expression close to background (Figure 1A). We then queried expression of myogenic genes (*Myod* and *Myog*) using more fully differentiated parental and cKO cells (D19) and RT-qPCR. Relative to Rps2 (which encodes a ribosomal 40S subunit), both genes were strongly expressed in parental but not cKO cells (Figure 1B). These results demonstrate that cKO ESCs differentiate into muscle poorly, consistent with their defective early gene expression programs and developmental trajectories.

To determine the effect of ectopic Oct1 expression, 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 to D11, and tested for *Myod*, *Myog* and *Pax3* by RT-qPCR. Differentiating cKO ESCs transduced with Oct1 but not empty vector expressed *Myod* and *Myog* more strongly. *Pax3* expression by contrast was reduced (Figure 1C). The combination of elevated *Myod* and *Myog* with diminished *Pax3* expression at late differentiation timepoints suggests that restoration of Oct1 allows cells to more efficiently transit through a *Pax3*-expressing intermediate state, such that more cells enter into a terminal myogenic program. Interestingly retrovirally expressed ectopic Oct1 expression also improved differentiation outcomes using parental cells (Figure 1C). Immunoblotting confirmed ectopic Oct1 expression and complementation of Oct1 deficiency (Figure 1D).

Defective lineage-appropriate gene expression differentiating Oct1-deficient cells

To identify how Oct1 loss perturbs MD differentiation, we performed bulk RNA-seq using parental and Oct1 cKO cells. Three replicates were performed per condition. ~1700, ~800 and 3,300 significantly differentially expressed genes were identified at D0, D3 and D6, respectively (FDR ≤ 0.05; -1 < log2FC < 1, Supplementary Table 1). Euclidean distance analysis reveals tight correlations between replicates and conditions at D0 and D3, but divergence at D6 relative to the other conditions, and devergence between parental and cKO D6 replicates (Supplemental Figure 1). Unsupervised hierarchical clustering reveals groups of genes regulated similarly between parental and cKO cells, and others with differential expression (Figure 1E). Cluster 2 for example shows strong induction in differentiating parental but failed induction in cKO D6 cells. Example genes in this set include *Pax3*, *Pax7* and *Myog* (Figure 1E, box). GO analysis of the differentially expressed genes in cluster 2 reveals association with differentiation to neural tube (neuroectoderm) and somite (Figure 1F). By contrast, clusters 1 and 4, which were associated with aberrant gene expression in differentiating cKO cells, were associated with lineage-inappropriate terms such as primordial germ cells, vascular system and stromal cells (Figure 1F and not shown). To identify potential regulators of cluster 2 genes with failed induction in differentiating cKO cells, we queried the ChIP Enrichment Analysis (ChEA) database (Lachmann et al., 2010). Cluster 2 genes tend to be bound by the polycomb repressor compex 2 (PRC2) components Suz12 and Mtf2 (Figure 1F). 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 with divergence at D6 (Supplemental Figure 1), as the 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.

Differentiating ESCs lacking Oct1 proceed down abnormal developmental trajectories

To investigate perturbations in cell populations and gene expression at single-cell resolution, we performed single-cell RNA-seq (scRNA-seq) using parental and cKO ESCs, and cells differentiated towards MD for 3 or 6 days. Cell numbers ranged between 1000 and 2500, with reads per cell ranging between 90,000 and 200,000 (see methods). Data from the different conditions were subjected to graph-based clustering and uniform manifold approximation and projection (UMAP) to visualize clusters of cells with similar gene expression. Integrated analysis of undifferentiated parental ESCs and parental cells at the two differentiation timepoints identifies

a range of expected cell populations, from naïve and primed pluripotency to neuromesodermal, paraxial MD and early somite (Figure 2A). 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 2B, Supplementary Figure 2 and not shown). These results provide a single cell-resolution map of early MD differentiation.

Next, we compared the parental controls to cKO cells to identify changes in cell populations and gene expression. D0 and D3 cKO cells show few differences from parental, Oct1 sufficient cells (Supplementary Table 2). In contrast, D6 cKO cells show poor differentiation capacity with reductions in key mesodermal clusters (Figure 2C). For example, populations characterized by neuromosodermal and paraxial MD gene expression represent the largest two parental cell clusters, accounting for 14.7 and 12.9% of cells, respectively, while populations associated with blood progenitors and somites account for 5.7% and 5.6% of cells (Figure 2C, left panel). The complexity in the populations is 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 clusters that retain epithelial characteristics and dramatically decreased percentages of neuromesodermal progenitors (2.3%, a six-fold decrease), paraxial MD (7.5%, two-fold), and both blood progenitors (1.1%, >5-fold) and somites (0.5%, >10-fold, Figure 2C, right panel). This failure to produce more differentiated MD lineage cells is consistent with findings that ESCs lacking Oct1 differentiate poorly into myotubes (Figure 1A).

Gene expression comparisons between parental and cKO paraxial mesoderm clusters reveals that cKO cells fail to robustly 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 2D). 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 3A, 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 2C

and Supplementary Figure 2). 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 the degree of expression for the indicated gene on the Y-axis (Figure 3B). 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 3B, 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 3C). This analysis revealed clusters associated with pluripotency (*Klf4*, *Nanog, Pou5f1*), early MD differentiation (*T*), paraxial mesoderm (*Rspo3, Meox1*), and early somites (*Dll1, Pax3*). The pseudotime gene expression pattern for cKO cells revealed largely normal expression of early gene expression markers, but loss of temporal coherence at subsequent steps. These "fuzzy" patterns of gene expression begin early in differentiation, with for example prolonged expression of genes such as *Tbxt* and noncoherent expression of genes associated with paraxial mesoderm such as *Meox1* and *Rspo3*. The diffuse pattern is different from what would be expected with a kinetic delay, which would cause a rightward shift. In addition, the induction of somitic genes normally expressed later in development such as *Dll1*, *Pax3* and *Pax7* was poor in cKO cells (Figure 3C).

RNA velocity analysis allows developmental directionality to be inferred relative to other cells by comparing 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. We clustered the cells separately by genotype, resulting in different patterns between genotype and compared to cells clustered together (Figure 3D). Applying velocity to this clustering, we found that D6 differentiated parental cells form discrete clusters with dynamic developmental progression, e.g. paraxial MD and somite (Figure 3D, 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 an aberrant epithelial state and multiple lineage markers (Figure 3D, epithelial clusters). Cumulatively, the data demonstrate that Oct1 is required for efficient early adoption of the MD differentiation program, with Oct1-deficient cells unable to canalize appropriate lineage specification programs.

241

242 **Oct1 occupies developmental genes in ESCs**

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

256 In pluripotent cells, ~22,000 Oct4 peaks were identified, corresponding to ~6,000 genes
257 with transcription start sites (TSS) within 20 kb (Supplementary Table 3). ~45% of Oct4 targets
258 directly overlap Oct1 peaks. Conversely ~60% of Oct1 targets overlap Oct4 peaks (Figure 4A),
259 indicating substantial overlap between binding of the two transcription factors. Shared Oct1/Oct4
260 targets in ESCs include *Polr2a*, which encodes the largest subunit of RNA polymerase II, *Pou5f1*,
261 which encodes Oct4, and *Dll1*, which encodes a developmentally-inducible mediator of Notch
262 signaling expressed in the MD lineage where it regulates muscle development (Zhang et al.,
263 2021). Undifferentiated (D0) Oct4 and Oct1 binding tracks are shown in Supplementary Figure 3.
264 We also confirmed Oct1 and Oct4 binding to *Polr2a*, *Pou5f1* and *Dll1* using ChIP-qPCR
265 (Supplementary Figure 3). Oct1 binding to *Polr2a*, which contains two adjacent MOREs that can
266 bind four Oct proteins (Kang et al., 2009a), is exceptional relative to other genes in that it is far
267 stronger than Oct4 (100× stronger for *Polr2a*, 3-10× weaker for *Pou5f1* and *Dll*). Re-ChIP
268 (sequential ChIP) indicates that Oct1 and Oct4 bind these sites simultaneously. The signal was
269 lost in Oct1 cKO ESCs, indicating specificity for Oct1 (Supplementary Figure 3). Cumulatively, the
270 data indicate that in ESCs Oct1 co-binds with Oct4 to an array of targets, including developmental-
271 specific targets.

272 We then performed ChIP-seq using D3 and D6 MD-differentiated cells. Only 199 Oct4-
273 bound peaks not occupied by Oct1 in pluripotent cells become occupied by Oct1 at D6 of MD

differentiation (Figure 4A). In contrast, 807 shared Oct4/Oct1 peaks 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 (Supplementary Figure 4A). Additionally, >6000 peaks become uniquely bound by Oct1 at MD D6 (Figure 4A). >2300 peaks are uniquely bound at D3 (Supplementary Figure 4B).

To pattern Oct1 and Oct4 occupancy during MD differentiation, we applied hierarchical clustering. Three major clusters of occupied targets were identified, two of which show decreased or static Oct1 binding over time (Figure 4B, clusters 1 and 2). Gene ontogeny (GO) analysis indicates involvement in signaling, mitochondrial function, stem cell differentiation and the regulation of Notch and TGFβ signaling. Oct1 and Oct4 enrichment at a cluster 1 gene encoding a pluripotency factor (*Pou5f1*) and two cluster 2 genes encoding MD/myogenic transcription factors (*Pax3* and *Myog*) is shown in Figure 4C. We also queried Oct1 binding to genes associated with bivalency, which can be used as a surrogate for silent, poised developmental genes. To identify bivalent genes, we intersected ESC H3K27me3- and H3K4me3-enriched ChIP-seq peaks from the ENCODE database. This procedure identifies 3861 bivalent genes (Supplementary Table 4). 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 4D, 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, Supplementary Table 5) shows even stronger Oct1 enrichment (Figure 4D, 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 during differentiation.

In contrast, Oct1 binding in Figure 4B increases with differentiation in cluster 3. GO analysis indicates that genes near these peaks are associated with chromatin-modifying activities including histone acetyltransferase and polycomb group complexes. Oct1 and Oct4 enrichment at example cluster 3 genes encoding polycomb complex members (*Ezh2*, *Suz12*, *Ring1* and *Ezh1*) is shown in Supplementary Figure 5A. These genes 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 (Supplementary Figure 5B-D). These data indicate that during MD differentiation, Oct1

directly binds to, and mediates the induction of, genes encoding epigenetic regulators associated with H3K27me3.

Oct1 Recruits Utx to the *Pax3* Target Gene to Remove H3K27me3. Lineage-appropriate poised developmental genes become activated in part via the removal of repressive H3K27me3 marks (Bernstein et al., 2006; Dhar et al., 2016a). Our findings indicate that genes with failed induction in differentiating cKO cells tend to be bound by chromatin-modifying complexes that act on H3K27me3 (Figure 1F). We hypothesized that Oct1 locally recruits H3K27-specific demethylases to lineage-specific genes to mediate their induction during differentiation. One such demethylase is Utx/Kdm6a (Cloos et al., 2008), which has been shown to regulate MD differentiation (Wang et al., 2012). To test the association between Oct1 binding and Utx, we also performed ChIP-seq using H3K27me3 and Utx 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 (Supplementary Table 3). 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 (Figure 5A, cluster 3), and strong Oct1 binding events associated with weaker Utx and lack of H3K27me3 (cluster 4). This latter cluster includes genes induced in MD such as *Pax3, Pax7* and *Myog*. We interpret cluster 3 to be developmental genes that remain poised at this timepoint, and cluster 4 to be lineage-appropritate genes that have been activated. 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, and reduced in Oct1 cKO cells at this timepoint (Figure 1E and Figure 3C). As expected, H3K27me3 was robust and equivalent in parental and cKO D0 undifferentiated cells. D6 parental cells showed reduced H3K27me3, while in contrast, cKO cells inappropriately retained elevated H3K27me3 (Figure 5D). This failure to remove H3K27me3 resulted in ~3-fold higher H3K27me3 enrichment 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). These results indicate 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 and Utx interact using extracts from undifferentiated and D6-differentiated parental ESCs and co-immunoprecipitation. The Oct1-Utx interaction is present even in ESCs, and is maintained during differentiation (Figure 5G). These results support the interpretation that Oct1 interacts with Utx and recruits it to target genes during differentiation to help mediate their activation. Utx recruitment by Oct1 potentially explains the failed induction of lineage-appropriate genes such as *Pax3*, and the failed removal of H3K27me3 in cKO cells (Figure 5D). Oct1, Oct4 and Utx binding to *Pax3* is shown in Figure 5H.

Oct1 interacts with Smad3, and cooperates with Smad proteins at *Myog* in cells differentiating towards MD. The broad expression of Oct1 raised the question of how Oct1 mediates gene activation specifically at MD lineage-specific. 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 4A). This procedure identifies not only Oct1 sites, but also binding sites for Smads, the terminal transcription regulators of TGFβ/Nodal/BMP 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 and 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 (Liu et al., 2011). Mammalian Smad3 also interacts with Oct4 and co-occupies target genes in pluripotent cells (Mullen et al., 2011). Consistent with these findings, Oct1 antibodies efficiently co-precipitate Smad3 in D6 MD-differentiated cells (Figure 6B). *Smad3* is expressed in undifferentiated parental populations, but further induced by MD differentiation (Figure 6C).

To gain more insight into mechanism, we tested if Oct1 and Smad proteins 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 into MEFs lacking Oct1 (Victoria E. H. Wang et al., 2004) (Figure 6D). Cells were either co-transfected with a construct encoding mouse Oct1, treated with recombinant TGFβ1, 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.

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 configuration that allows for later induction or stable repression, depending on the developmental lineage (Bernstein et al., 2006). Oct1-deficient animals manifest defective induction of MD-derived somites, cardiac tissue and blood cells (Sebastiano et al., 2010; V. E. H. Wang et al., 2004). The co-expression of Oct1 with Oct4 in pluripotent cells and maintenance of Oct1 beyond Oct4 silencing (Shen et al., 2017), together with the severe developmental phenotype of Oct1 null animals, prompted us to study the role of Oct1 during differentiation into the MD lineage. Here we show that Oct1-deficient ESCs manifest defective MD differentiation, including failure to express *Myod* and *Myog* mRNA, and myosin heavy chain protein. During mesodermal differentiation, Oct1 continues to bind developmental lineage-specific genes co-bound with its paralog Oct4 in pluripotent cells. Further, Oct1 helps mediate the induction of the subset of MD lineage-specific. We show that Oct1 and Utx binding events significantly overlap, that Oct1 interacts with Utx and that Oct1 localizes Utx to a MD-specific target to remove repressive H3K27me3 marks. Specificity is achieved at least in part through cooperative interactions with Smad proteins, known drivers of MD induction. Because Oct1 is widely expressed and results in defective differentiation into multiple lineages (Shen et al., 2017), we anticipate that this mechanism may act widely, with the only unique aspect being the association with lineage-driving transcription regulators such as Smad proteins in the case of MD.

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 an epithelial state, and achieve neuromesodermal and somite-stage gene expression patterns poorly and in reduced numbers. Increased populations of cells appear associated with oxidative stress, consistent with findings that cells lacking Oct1 have elevated reactive oxygen species levels (Tantin et al., 2005). cKO cells mis-express developmentally inappropriate genes and undergo inappropriate developmental branching towards poorly differentiated states marked by epithelial gene expression and oxidative stress. Unlike the expected effect of loss of lineage-specific master regulators where cell fate often shifts to a specific alternate lineage, these cells lack clear lineage signatures. 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 MD-differentiating cells, Oct1 occupies a subset of genes bound by Oct4 in pluripotent cells. These include lineage-specific developmental mediators and genes encoding chromatin-modifying enzymes. Oct1 occupancy on these genes 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 promoter and the failed transcriptional upregulation. No global changes in H3K27me3 were observed, consistent with recruitment by Oct1 to mediate focal removal (Wang et al., 2012). 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). To our knowledge this is the first description of a role for Oct1 and K3K27me3. Loss of Utx allows for maintainance of an undifferentiated state, but interferes with MD differentiation (Dhar et al., 2016b).

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

Under MD differentiation conditions, exogenous Oct1 increases expression of the terminal differentiation markers *Myod* and *Myog*, while decreasing the early lineage-specification marker *Pax3*, which is transiently expressed during MD differentiation and is necessary for later expression of myogenic genes. Because *Pax3* is no longer expressed in terminally differentiating cells, these results suggest that ectopic Oct1 may enable transit through a *Pax3*-expressing intermediate to potentiate productive terminal differentiation. More investigation into this pathway

445 is necessary to test if Oct1 can be used to more efficiently differentiate pluripotent cells. Such
446 improvements could have implications for regenerative medicine.

Figure Legends

Figure 1. Failed MD/muscle lineage induction in differentiating Oct1 cKO ESCs. (**A**) 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 D11. (**B**) Relative mRNA expression levels of the myogenic genes *Myod1* and *Myog* in parental and cKO ESCs differentiated for 19D. Data represent an average of biological triplicates. Error bars depict ±SEM. (**C**) RT-qPCR for the myogenic genes *Myod1*, *Myog* and *Pax3* in parental and cKO cells transduced with retroviruses encoding Oct1 or empty vector. Transduced cells were selected using puromycin for 48 hr prior to differentiation for 11D. Empty vector values were set to 1. Data represent an average of biological triplicates. Error bars depict ±SEM. (**D**) Immunoblot showing ectopic Oct1 expression in parental and cKO cells. β-actin is shown as a loading standard. (**E**) 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 cluster 2 genes are shown at right. (**F**) Jensen Tissue and ChIP-X Enrichment Analysis (ChEA) query results for Cluster 2 in (B) are shown. (**G**) 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 2. 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). MD-associated clusters are shown in bold. 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. MD lineage-associated clusters are shown in bold. (**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 3. 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. Clusters in bold are associated with MD differentiation.

Figure 4. 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**) 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. (**C**) Example cluster 1 and 2 genome tracks: *Pou5f1* (cluster 1), *Pax3* (cluster 2) and *Myog* (cluster 2). (**D**) 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 5. Oct1 recruits Utx to demethylate H3K27me3 at lineage-appropriate genes. (**A**) Oct1, H3K27me3 and Utx ChIP-seq enrichment 2 kb around peak centers was computed for the merged peak list and shown as heatmap. (**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 $\geq$ 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 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 a nonspecific 28S region is shown. Data represent an average of 3 independent biological replicates. Error bars depict ±SEM. (**G**) Utx immunoprecipitation followed by Oct1 immunoblot using parental ESCs, or ESCs MD-differentiated for 6D. (**H**) 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 motifs are shown below.

Figure 6. Oct1 and Smad3 cooperate to drive expression of mesoderm-specific genes. (**A**) HOMER motif analysis of Oct1 peaks that are 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**) Scheme for Oct1/Smad reporter assay. A segment of a mouse *Myog* enhancer element containing multiple Oct 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. Secreted luciferase activity was assessed relative to secreted mCherry expressed from a co-transfected plasmid. An average of experimental triplicates 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. ENCODE-annotated regulatory elements are shown below. Pink: proximal promoter. Yellow: distal enhancer.

Figure 7. 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 co-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.

Methods

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

Cell culture. ESCs were cultured as previously described (Shakya et al., 2015) 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 Oct1 deletion, cells were treated with 4-hydroxytamoxifen and a completely yellow sub-colony was picked and expanded as published (Shen et al., 2017). 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).

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.

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 Supplementary Table 6 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.

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.

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.

Single-cell RNA-seq 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 (Figure 2A) or parental and cKO D6 (Figure 2C) 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. Using this analysis, the cell numbers were between 1000 and 2500. Reads/cell ranged between 90,000 and 200,000. Genes/cell ranged between 2,300 and 7,000. The percentage of reads mapping to

reference mouse genome was 90% or higher, and the reads confidently mapping to exonic regions was 70% or higher.

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 (Figure 2A). 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.

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 the primary antibody (either Utx or Oct1) 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 μL IP Elution buffer at RT, then diluted 10-fold in RIPA buffer. Diluted samples were then subjected to a second IP with 4 μg of the secondary antibody (Oct1 or Oct4) 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 μL 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 Supplementary Table 6 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 'chrMlPhiX' --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.

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

712 single or combined mutations in the Oct1 or Smad binding sites, fused upstream of the CMV basal
713 promoter (-216-13 relative to TSS). G-blocks were inserted using sequence- and ligase-
714 independent cloning (Li and Elledge, 2012) upstream of the coding sequence for a secreted nano-
715 luciferase following digestion of vector pNL2.3 (Promega) using *Eco*RV and *Hind*III. The veracity
716 of the cloned inserts was confirmed by Sanger sequencing. 200 ng of reporter plasmid were co-
717 transfected into Oct1-deficient MEFs (Shakya et al., 2009) in DMEM media lacking phenol red
718 (ThermoFisher) together with 400 ng MMP9-mCherry (Wider and Picard, 2017) in 1000 ng total
719 transfected DNA. Where indicated, 400 ng pBabePuro-Oct1 was included in the transfection mix.
720 pUC18 plasmid comprised the balance of the transfected DNA. Where indicated, transfected cells
721 were provided with 5 ng recombinant mouse TGFb1 protein (R&D Systems). mCherry
722 fluorescence was determined first by exciting at 570 nm and measuring emission at 610 nm with
723 a 100 msec time delay using an Envision Xcite Multilabel Plate Reader. Luminescence was
724 measured using Nano-Glo Luciferase (Promega) and a Modulus luminescence plate reader.
725
726 **Metabolic profiling.** Cold 90% methanol (MeOH) was added to each sample to give a final
727 concentration of 80%. Samples were then incubated at -20˚C for 1 hr. After incubation the
728 samples were centrifuged at 20,000×*g* for 10 min at 4˚C. The supernatant was then transferred
729 from each sample tube into a labeled, fresh micro centrifuge tube. Pooled quality control samples
730 were made by removing a fraction of collected supernatant from each sample and process blanks
731 were made using only extraction solvent and no cell culture. The samples were then dried en
732 vacuo. GC-MS was performed with an Agilent 5977b GC-MS MSD-HES and an Agilent 7693A
733 automatic liquid sampler. Data were analyzed using in-house software to prepare for analysis by
734 the "MetaboAnalyst" software tool (Chong et al., 2018). Statistical analysis was performed using
735 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.

Conflict of interest

The authors declare that they have no conflicts of interest

752 **REFERENCES**

753

754 Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, Aken B, Akiyama JA, Jammal O Al,
755 Amrhein H, Anderson SM, Andrews GR, Antoshechkin I, Ardlie KG, Armstrong J, Astley M,
756 Banerjee B, Barkal AA, Barnes IHA, Barozzi I, Barrell D, Barson G, Bates D, Baymuradov
757 UK, Bazile C, Beer MA, Beik S, Bender MA, Bennett R, Bouvrette LPB, Bernstein BE,
758 Berry A, Bhaskar A, Bignell A, Blue SM, Bodine DM, Boix C, Boley N, Borrman T, Borsari
759 B, Boyle AP, Brandsmeier LA, Breschi A, Bresnick EH, Brooks JA, Buckley M, Burge CB,
760 Byron R, Cahill E, Cai L, Cao L, Carty M, Castanon RG, Castillo A, Chaib H, Chan ET,
761 Chee DR, Chee S, Chen Hao, Chen Huaming, Chen JY, Chen S, Cherry JM, Chhetri SB,
762 Choudhary JS, Chrast J, Chung D, Clarke D, Cody NAL, Coppola CJ, Coursen J, D'Ippolito
763 AM, Dalton S, Danyko C, Davidson C, Davila-Velderrain J, Davis CA, Dekker J, Deran A,
764 DeSalvo G, Despacio-Reyes G, Dewey CN, Dickel DE, Diegel M, Diekhans M, Dileep V,
765 Ding B, Djebali S, Dobin A, Dominguez D, Donaldson S, Drenkow J, Dreszer TR, Drier Y,
766 Duff MO, Dunn D, Eastman C, Ecker JR, Edwards MD, El-Ali N, Elhajjajy SI, Elkins K, Emili
767 A, Epstein CB, Evans RC, Ezkurdia I, Fan K, Farnham PJ, Farrell NP, Feingold EA,
768 Ferreira AM, Fisher-Aylor K, Fitzgerald S, Flicek P, Foo CS, Fortier K, Frankish A, Freese
769 P, Fu S, Fu XD, Fu Y, Fukuda-Yuzawa Y, Fulciniti M, Funnell APW, Gabdank I, Galeev T,
770 Gao M, Giron CG, Garvin TH, Gelboin-Burkhart CA, Georgolopoulos G, Gerstein MB,
771 Giardine BM, Gifford DK, Gilbert DM, Gilchrist DA, Gillespie S, Gingeras TR, Gong P,
772 Gonzalez A, Gonzalez JM, Good P, Goren A, Gorkin DU, Graveley BR, Gray M, Greenblatt
773 JF, Griffiths E, Groudine MT, Grubert F, Gu M, Guigó R, Guo H, Guo Yu, Guo Yuchun,
774 Gursoy G, Gutierrez-Arcelus M, Halow J, Hardison RC, Hardy M, Hariharan M, Harmanci
775 A, Harrington A, Harrow JL, Hashimoto TB, Hasz RD, Hatan M, Haugen E, Hayes JE, He
776 P, He Y, Heidari N, Hendrickson D, Heuston EF, Hilton JA, Hitz BC, Hochman A, Holgren
777 C, Hou L, Hou S, Hsiao YHE, Hsu S, Huang H, Hubbard TJ, Huey J, Hughes TR, Hunt T,
778 Ibarrientos S, Issner R, Iwata M, Izuogu O, Jaakkola T, Jameel N, Jansen C, Jiang L, Jiang
779 P, Johnson A, Johnson R, Jungreis I, Kadaba M, Kasowski M, Kasparian M, Kato M, Kaul
780 R, Kawli T, Kay M, Keen JC, Keles S, Keller CA, Kelley D, Kellis M, Kheradpour P, Kim
781 DS, Kirilusha A, Klein RJ, Knoechel B, Kuan S, Kulik MJ, Kumar S, Kundaje A, Kutyavin T,
782 Lagarde J, Lajoie BR, Lambert NJ, Lazar J, Lee AY, Lee D, Lee E, Lee JW, Lee K, Leslie
783 CS, Levy S, Li B, Li H, Li N, Li X, Li YI, Li Ying, Li Yining, Li Yue, Lian J, Libbrecht MW, Lin
784 S, Lin Y, Liu D, Liu J, Liu P, Liu T, Liu XS, Liu Yan, Liu Yaping, Long M, Lou S, Loveland J,

785 Lu A, Lu Y, Lécuyer E, Ma L, Mackiewicz M, Mannion BJ, Mannstadt M, Manthravadi D,
786 Marinov GK, Martin FJ, Mattei E, McCue K, McEown M, McVicker G, Meadows SK,
787 Meissner A, Mendenhall EM, Messer CL, Meuleman W, Meyer C, Miller S, Milton MG,
788 Mishra T, Moore DE, Moore HM, Moore JE, Moore SH, Moran J, Mortazavi A, Mudge JM,
789 Munshi N, Murad R, Myers RM, Nandakumar V, Nandi P, Narasimha AM, Narayanan AK,
790 Naughton H, Navarro FCP, Navas P, Nazarovs J, Nelson J, Neph S, Neri FJ, Nery JR,
791 Nesmith AR, Newberry JS, Newberry KM, Ngo V, Nguyen R, Nguyen TB, Nguyen T,
792 Nishida A, Noble WS, Novak CS, Novoa EM, Nuñez B, O'Donnell CW, Olson S, Onate KC,
793 Otterman E, Ozadam H, Pagan M, Palden T, Pan X, Park Y, Partridge EC, Paten B, Pauli-
794 Behn F, Pazin MJ, Pei B, Pennacchio LA, Perez AR, Perry EH, Pervouchine DD, Phalke
795 NN, Pham Q, Phanstiel DH, Plajzer-Frick I, Pratt GA, Pratt HE, Preissl S, Pritchard JK,
796 Pritykin Y, Purcaro MJ, Qin Q, Quinones-Valdez G, Rabano I, Radovani E, Raj A,
797 Rajagopal N, Ram O, Ramirez L, Ramirez RN, Rausch D, Raychaudhuri S, Raymond J,
798 Razavi R, Reddy TE, Reimonn TM, Ren B, Reymond A, Reynolds A, Rhie SK, Rinn J,
799 Rivera M, Rivera-Mulia JC, Roberts BS, Rodriguez JM, Rozowsky J, Ryan R, Rynes E,
800 Salins DN, Sandstrom R, Sasaki T, Sathe S, Savic D, Scavelli A, Scheiman J, Schlaffner
801 C, Schloss JA, Schmitges FW, See LH, Sethi A, Setty M, Shafer A, Shan S, Sharon E,
802 Shen Q, Shen Y, Sherwood RI, Shi M, Shin S, Shoresh N, Siebenthall K, Sisu C, Slifer T,
803 Sloan CA, Smith A, Snetkova V, Snyder MP, Spacek D V., Srinivasan S, Srivas R,
804 Stamatoyannopoulos G, Stamatoyannopoulos JA, Stanton R, Steffan D, Stehling-Sun S,
805 Strattan JS, Su A, Sundararaman B, Suner MM, Syed T, Szynkarek M, Tanaka FY, Tenen
806 D, Teng M, Thomas JA, Toffey D, Tress ML, Trout DE, Trynka G, Tsuji J, Upchurch SA,
807 Ursu O, Uszczynska-Ratajczak B, Uziel MC, Valencia A, Biber B Van, van der Velde AG,
808 Van Nostrand EL, Vaydylevich Y, Vazquez J, Victorsen A, Vielmetter J, Vierstra J, Visel A,
809 Vlasova A, Vockley CM, Volpi S, Vong S, Wang H, Wang M, Wang Q, Wang R, Wang T,
810 Wang W, Wang X, Wang Y, Watson NK, Wei X, Wei Z, Weisser H, Weissman SM, Welch
811 R, Welikson RE, Weng Z, Westra HJ, Whitaker JW, White C, White KP, Wildberg A,
812 Williams BA, Wine D, Witt HN, Wold B, Wolf M, Wright J, Xiao R, Xiao X, Xu J, Yan KK,
813 Yan Y, Yang H, Yang X, Yang YW, Yardimci GG, Yee BA, Yeo GW, Young T, Yu T, Yue F,
814 Zaleski C, Zang C, Zeng H, Zeng W, Zerbino DR, Zhai J, Zhan L, Zhan Y, Zhang B, Zhang
815 Jialing, Zhang Jing, Zhang K, Zhang L, Zhang P, Zhang Q, Zhang XO, Zhang Y, Zhang Z,
816 Zhao Y, Zheng Y, Zhong G, Zhou XQ, Zhu Y, Zimmerman J. 2020. Expanded
817 encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**:699–710.

doi:10.1038/s41586-020-2493-4

Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI. 1995. Embryonic stem cells express neuronal properties in vitro. *Dev Biol* **168**:342–357. doi:10.1006/dbio.1995.1085

Beddington RSP, Robertson EJ. 1989. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**:733–737.

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**:315–26. doi:10.1016/j.cell.2006.02.041

Chal J, Oginuma M, Tanoury Z Al, Gobert B, Sumara O, Hick A, Bousson F, Zidouni Y, Mursch C, Moncuquet P, Tassy O, Vincent S, Miyanari A, Bera A, Garnier J, Guevara G, Hestin M, Kennedy L, Hayashi S, Drayton B, Cherrier T, Gayraud-morel B, Gussoni E, Relaix F, Tajbakhsh S, Pourquié O. 2015. Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nat Biotechnol* **33**:962–9. doi:10.1038/nbt.3297

Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. 2018. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* **46**:W486–W494. doi:10.1093/nar/gky310

Cloos PAC, Christensen J, Agger K, Helin K. 2008. Erasing the methyl mark: Histone demethylases at the center of cellular differentiation and disease. *Genes Dev*. doi:10.1101/gad.1652908

Collinson A, Collier AJ, Morgan NP, Sienerth AR, Chandra T, Andrews S, Rugg-Gunn PJ. 2016. Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. *Cell Rep*. doi:10.1016/j.celrep.2016.11.032

Conlon FL, Lyons KM, Takaesu N, Barth KS, Kispert A, Herrmann B, Robertson EJ. 1994. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Trends Genet* **10**:308–308. doi:10.1016/0168-9525(94)90026-4

Dell'Orso S, Juan AH, Ko K-D, Naz F, Perovanovic J, Gutierrez-Cruz G, Feng X, Sartorelli V. 2019. Correction: Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and regenerative conditions (doi: 10.1242/dev.174177). *Development* **146**:dev181743. doi:10.1242/dev.181743

DeVeale B, Brokhman I, Mohseni P, Babak T, Yoon C, Lin A, Onishi K, Tomilin A, Pevny L, Zandstra PW, Nagy A, van der Kooy D. 2013. Oct4 is required ~E7.5 for proliferation in the

primitive streak. *PLoS Genet* **9**. doi:10.1371/JOURNAL.PGEN.1003957

Dhar SS, Lee SH, Chen K, Zhu G, Oh WK, Allton K, Gafni O, Kim YZ, Tomoiga AS, Barton MC, Hanna JH, Wang Z, Li W, Lee MG. 2016a. An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic Acids Res* **44**:3659–3674. doi:10.1093/nar/gkv1516

Dhar SS, Lee SH, Chen K, Zhu G, Oh WK, Allton K, Gafni O, Kim YZ, Tomoiga AS, Barton MC, Hanna JH, Wang Z, Li W, Lee MG. 2016b. An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic Acids Res* **44**:3659–3674. doi:10.1093/nar/gkv1516

Diaz-Cuadros M, Wagner DE, Budjan C, Hubaud A, Tarazona OA, Donelly S, Michaut A, Al Tanoury Z, Yoshioka-Kobayashi K, Niino Y, Kageyama R, Miyawaki A, Touboul J, Pourquié O. 2020. In vitro characterization of the human segmentation clock. *Nature* **580**:113–118. doi:10.1038/s41586-019-1885-9

Gnocchi VF, White RB, Ono Y, Ellis JA, Zammit PS. 2009. Further characterisation of the molecular signature of quiescent and activated mouse muscle satellite cells. *PLoS One* **4**:e5205. doi:10.1371/journal.pone.0005205

Guibentif C, Griffiths JA, Imaz-Rosshandler I, Ghazanfar S, Nichols J, Wilson V, Göttgens B, Marioni JC. 2021. Diverse Routes toward Early Somites in the Mouse Embryo. *Dev Cell* **56**:141-153.e6. doi:10.1016/j.devcel.2020.11.013

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**:576–589. doi:10.1016/j.molcel.2010.05.004

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. 2013. Super-enhancers in the control of cell identity and disease. *Cell* **155**:934–47. doi:10.1016/j.cell.2013.09.053

Kang J, Gemberling M, Nakamura M, Whitby FG, Handa H, Fairbrother WG, 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. doi:10.1101/gad.1750709

Kang J, Gemberling M, Nakamura M, Whitby FG, Handa H, Fairbrother WG, 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. doi:10.1101/gad.1750709

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 Genet* **4**:e1000242. doi:10.1371/journal.pgen.1000242

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko P V. 2018. RNA velocity of single cells. *Nature* **560**:494–498. doi:10.1038/s41586-018-0414-6

Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. 2010. ChEA: Transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**:2438–2444. doi:10.1093/bioinformatics/btq466

Li MZ, Elledge SJ. 2012. SLIC: A method for sequence- and ligation-independent cloning. *Methods Mol Biol* **852**:51–59. doi:10.1007/978-1-61779-564-0_5

Liu Z, Lin X, Cai Z, Zhang Z, Han C, Jia S, Meng A, 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. doi:10.1074/jbc.M111.236307

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**:495–501. doi:10.1038/nbt.1630

Meshorer E, Misteli T. 2006. Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* **7**:540–6. doi:10.1038/nrm1938

Mullen AC, Orlando DA, Newman JJ, Lovén J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, DeKoter RP, Young RA. 2011. Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell* **147**:565–76. doi:10.1016/j.cell.2011.08.050

Nichols J, Zevnik B, Anastassiadis 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

Oginuma M, Moncuquet P, Xiong F, Karoly E, Chal J, Guevorkian K, 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. doi:10.1016/j.devcel.2017.02.001

Palmieri SL, Peter W, Hess H, Schöler HR. 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. doi:10.1006/dbio.1994.1312

Perovanovic J, DellOrso S, Gnochi VF, Jaiswal JK, Sartorelli V, Vigouroux C, Mamchaoui K, Mouly V, Bonne G, Hoffman EP. 2016. Laminopathies disrupt epigenomic developmental programs and cell fate. *Sci Transl Med* **8**:335ra58-335ra58. doi:10.1126/scitranslmed.aad4991

Schlesinger S, Meshorer E. 2019. Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. *Dev Cell*. doi:10.1016/j.devcel.2019.01.003

Sebastiano V, Dalvai M, Gentile L, Schubart K, Sutter J, Wu GM, Tapia N, Esch D, Ju JY, Hübner K, Bravo MJA, Schöler HR, Cavaleri F, Matthias P. 2010. Oct1 regulates trophoblast development during early mouse embryogenesis. *Development* **137**:3551–3560. doi:10.1242/dev.047027

Seydoux G, Braun RE. 2006. Pathway to Totipotency: Lessons from Germ Cells. *Cell*. doi:10.1016/j.cell.2006.11.016

Shakya A, Callister C, Goren A, Yosef N, Garg N, Khoddami V, Nix D, Regev A, Tantin D. 2015. Pluripotency transcription factor Oct4 mediates stepwise nucleosome demethylation and depletion. *Mol Cell Biol* **35**:1014–25. doi:10.1128/MCB.01105-14

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. *Nat Cell Biol* **11**:320–327. doi:10.1038/ncb1840

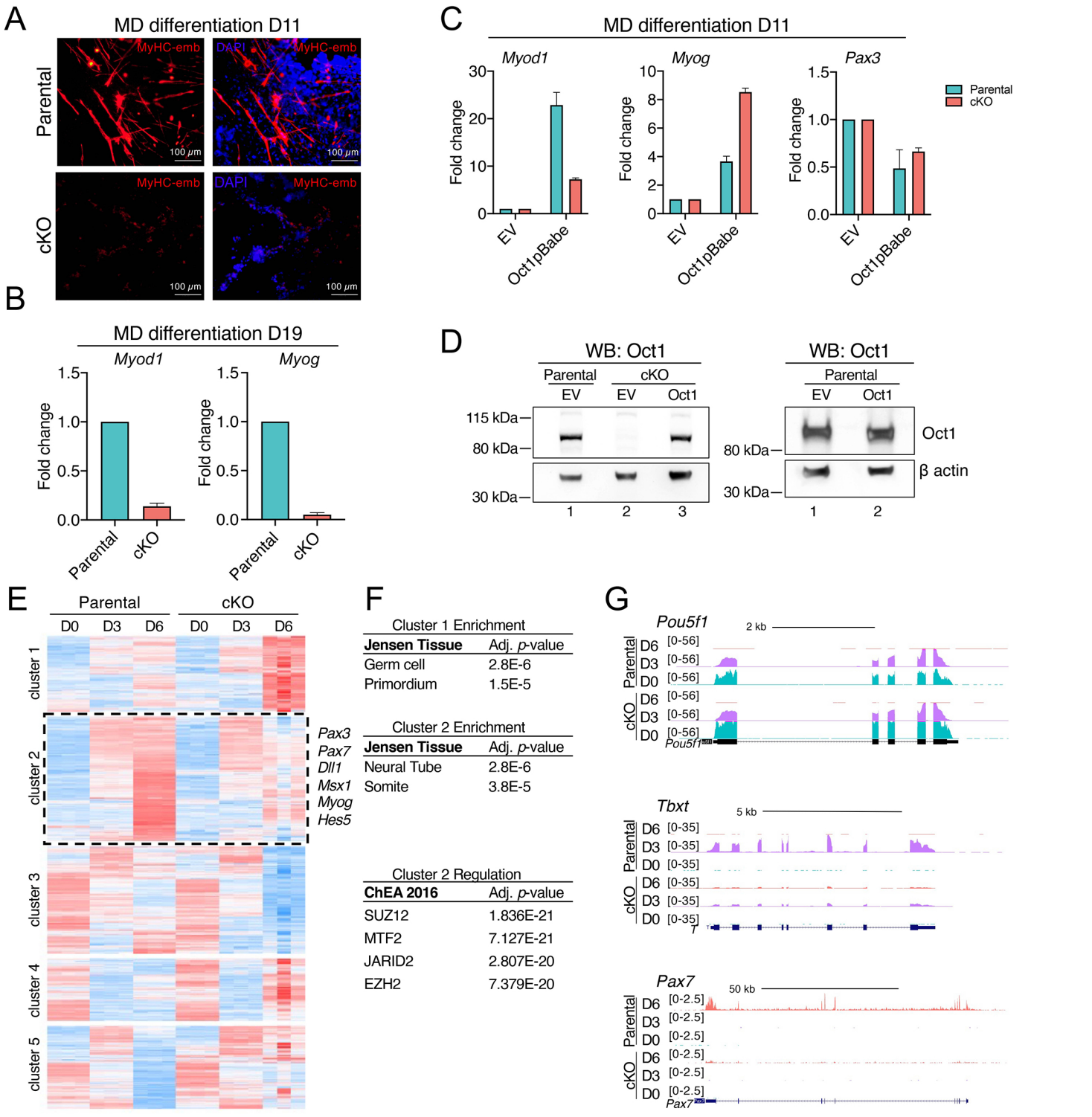
Shakya A, Kang J, Chumley J, Williams MA, Tantin D. 2011. Oct1 is a switchable, bipotential stabilizer of repressed and inducible transcriptional states. *J Biol Chem* **286**:450–9. doi:10.1074/jbc.M110.174045

Shen Z, Kang J, Shakya A, Tabaka M, Jarboe EA, Regev A, Tantin D. 2017. Enforcement of developmental lineage specificity by transcription factor Oct1. *Elife* **6**. doi:10.7554/eLife.20937

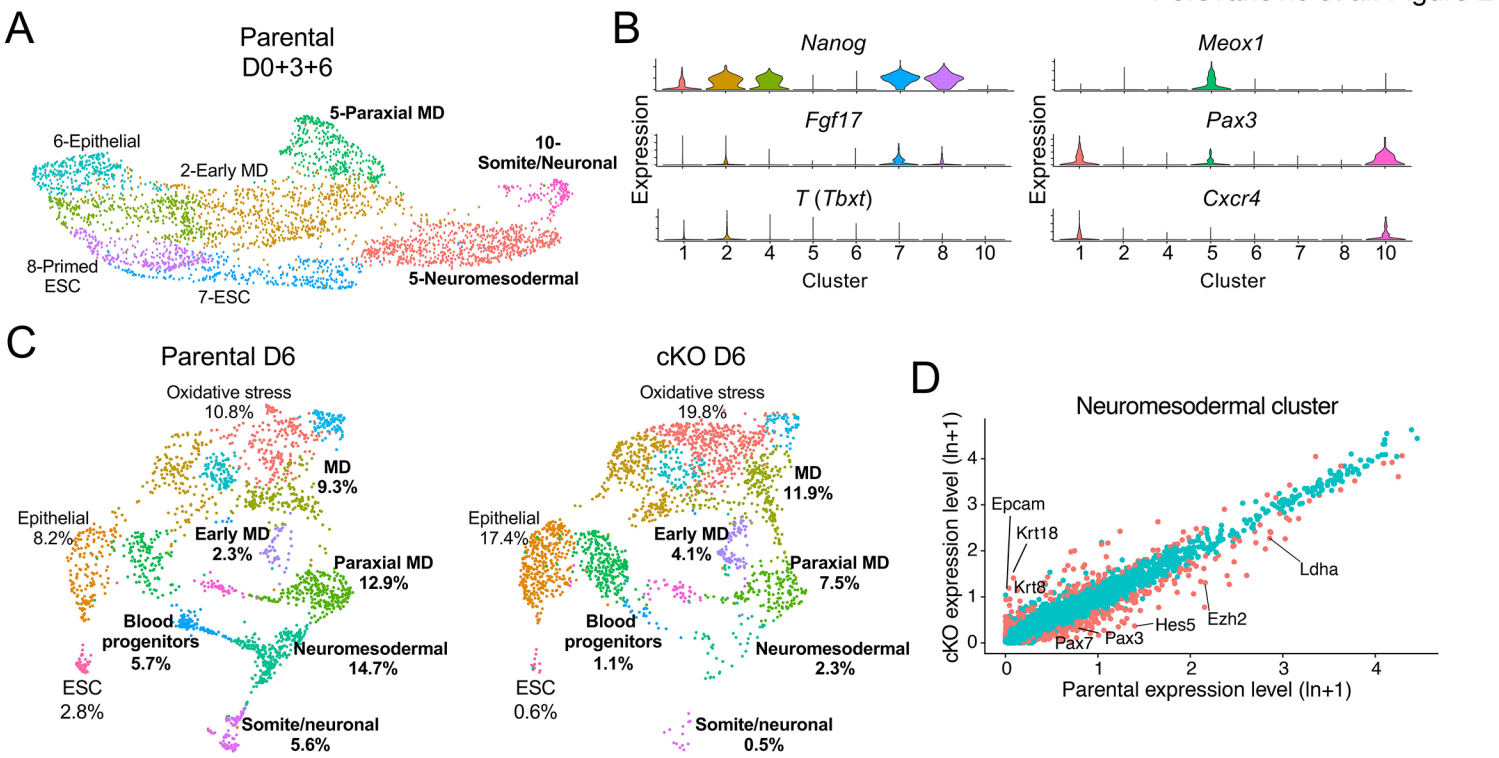
Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-1902.e21. doi:10.1016/j.cell.2019.05.031

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

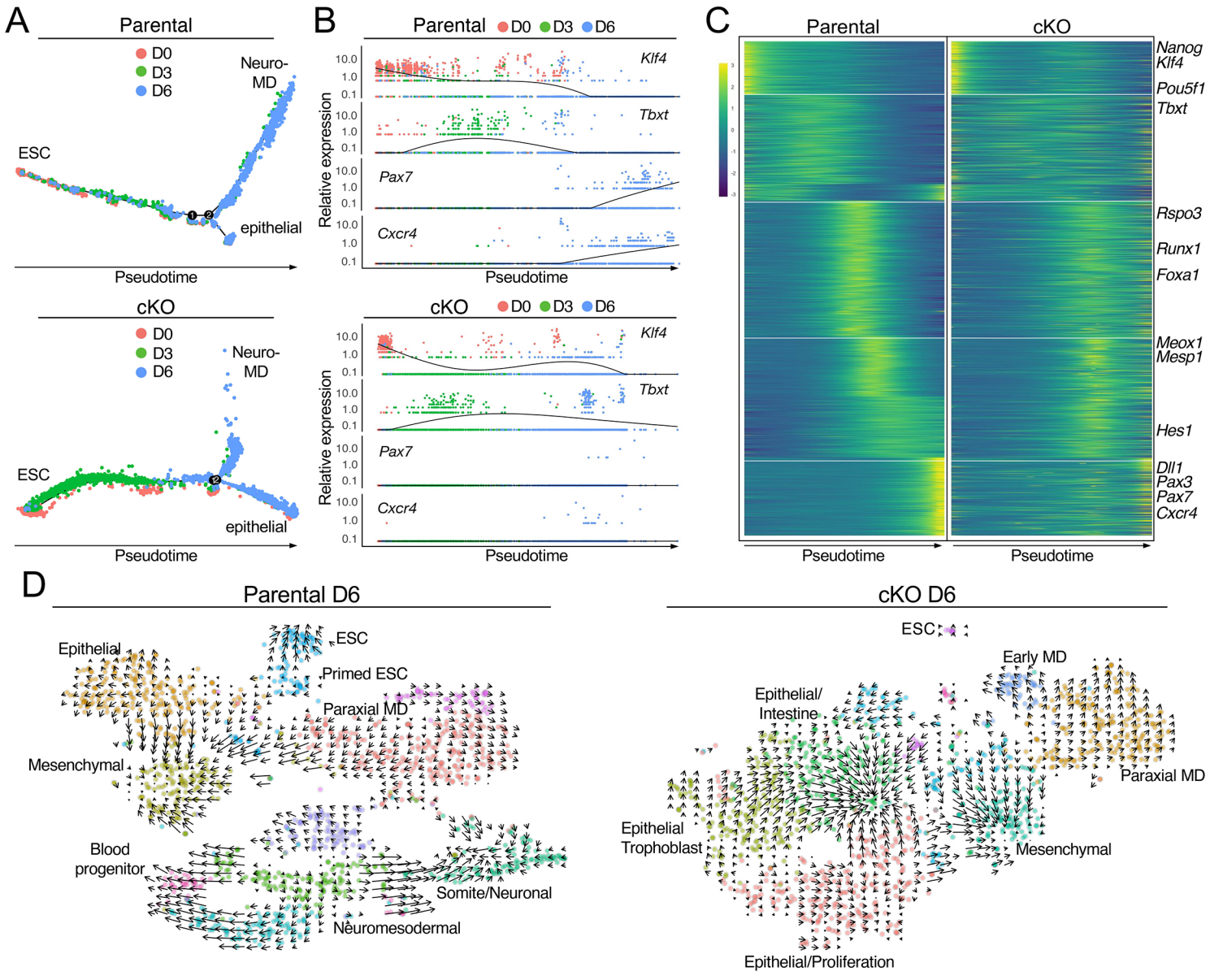
950 Tantin D. 2013. Oct transcription factors in development and stem cells: insights and
951 mechanisms. *Development* **140**:2857–2866. doi:10.1242/dev.095927
952 Tantin D, Schild-Poulter C, Wang V, Haché RJG, Sharp PA. 2005. The Octamer Binding
953 Transcription Factor Oct-1 Is a Stress Sensor. *Cancer Res* **65**:10750–10758.
954 doi:10.1158/0008-5472.CAN-05-2399
955 Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ,
956 Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are
957 revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381–386.
958 doi:10.1038/nbt.2859
959 Wang C, Lee JE, Cho YW, Xiao Y, Jin Q, Liu C, Ge K. 2012. UTX regulates mesoderm
960 differentiation of embryonic stem cells independent of H3K27 demethylase activity. *Proc*
961 *Natl Acad Sci U S A* **109**:15324–15329. doi:10.1073/pnas.1204166109
962 Wang V. E. H., Schmidt T, Chen J, Sharp PA, Tantin D. 2004. Embryonic Lethality, Decreased
963 Erythropoiesis, and Defective Octamer-Dependent Promoter Activation in Oct-1-Deficient
964 Mice. *Mol Cell Biol* **24**:1022–1032. doi:10.1128/mcb.24.3.1022-1032.2004
965 Wang Victoria E. H., Schmidt T, Chen J, Sharp PA, Tantin D. 2004. Embryonic lethality,
966 decreased erythropoiesis, and defective octamer-dependent promoter activation in Oct-1-
967 deficient mice. *Mol Cell Biol* **24**:1022–1032. doi:10.1128/MCB.24.3.1022-1032.2004
968 Wider D, Picard D. 2017. Secreted dual reporter assay with Gaussia luciferase and the red
969 fluorescent protein mCherry. *PLoS One* **12**:e0189403. doi:10.1371/journal.pone.0189403
970 Zhang Y, Lahmann I, Baum K, Shimojo H, Mourikis P, Wolf J, Kageyama R, Birchmeier C.
971 2021. Oscillations of Delta-like1 regulate the balance between differentiation and
972 maintenance of muscle stem cells. *Nat Commun* **12**. doi:10.1038/s41467-021-21631-4
973 Zhou X, Sasaki H, Lowe L, Hogan BLM, Kuehn MR. 1993. Nodal is a novel TGF-β-like gene
974 expressed in the mouse node during gastrulation. *Nature* **361**:543–547.
975 doi:10.1038/361543a0
976



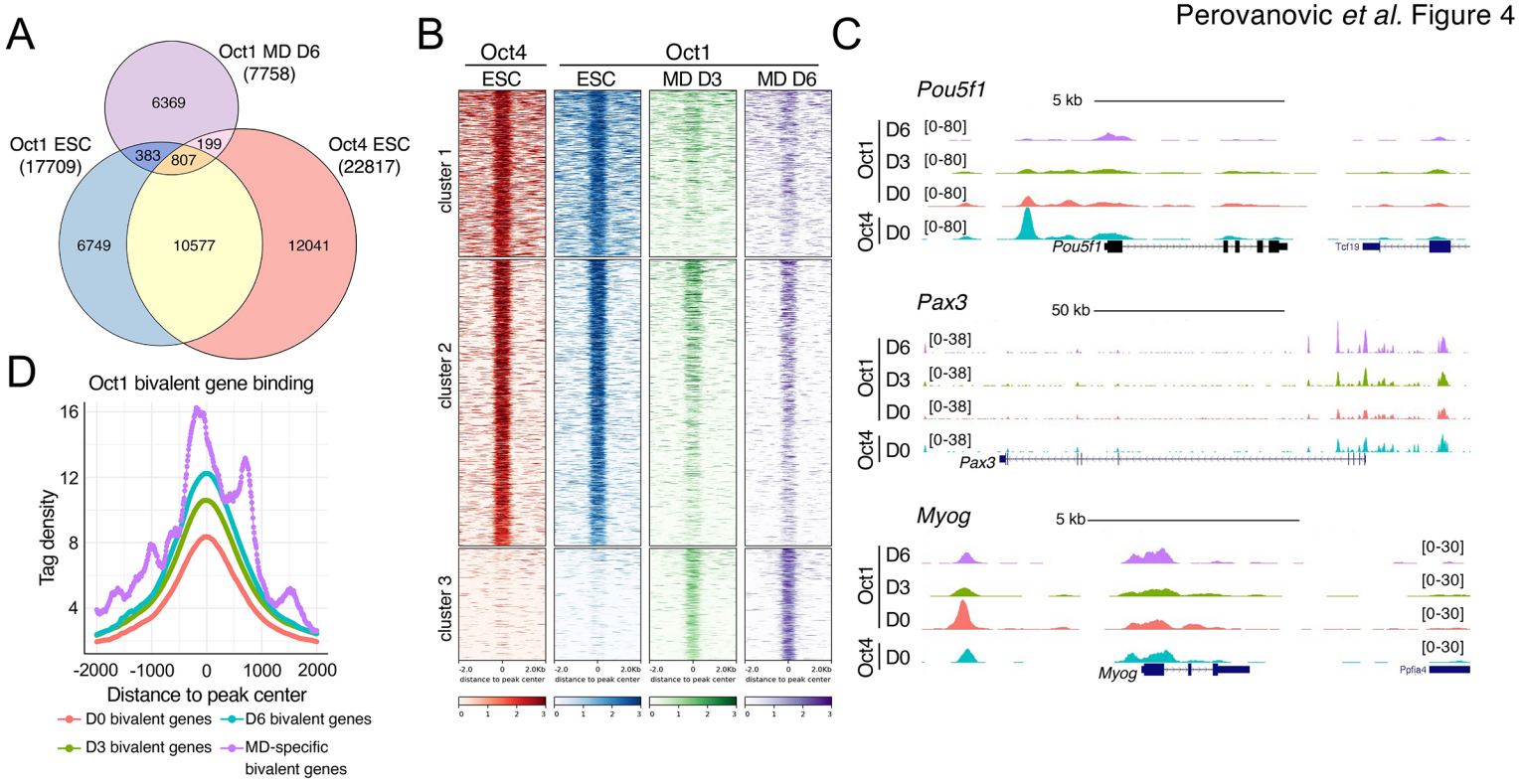
Perovanovic *et al.* Figure 2

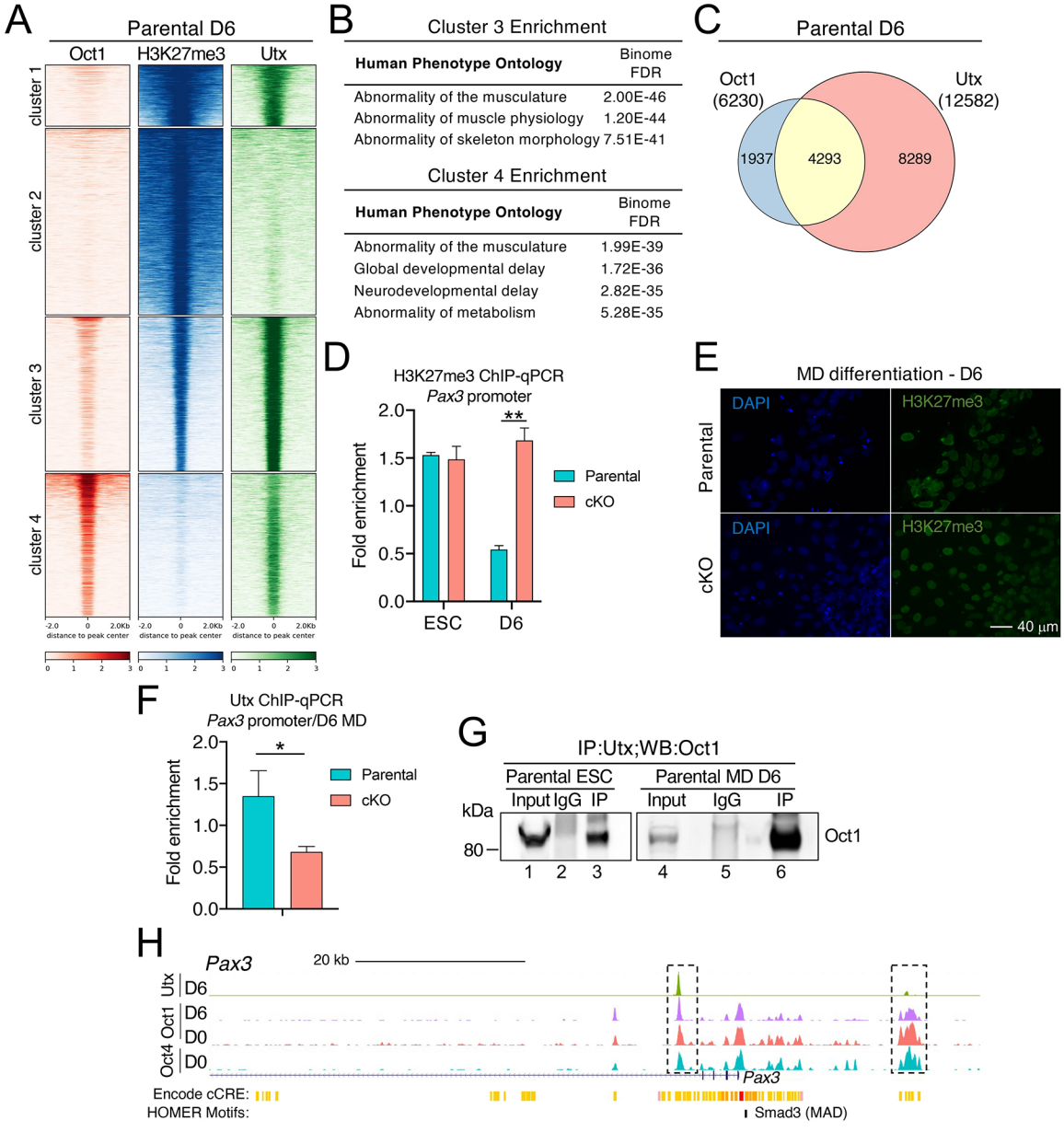


Perovanovic *et al.* Figure 3

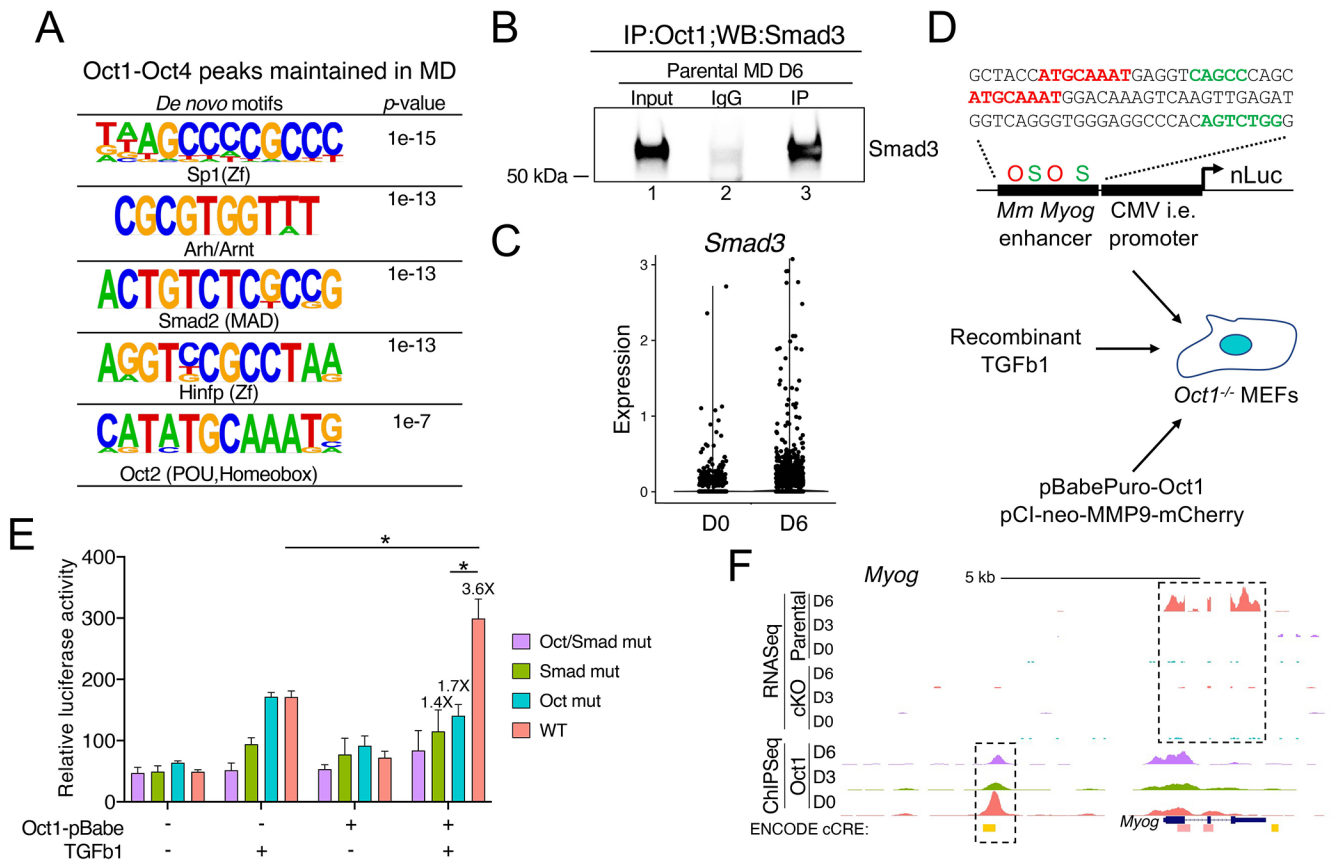


Perovanovic *et al.* Figure 4





Perovanovic *et al.* Figure 6



Perovanovic *et al.* Figure 7