Oct1 recruits the histone lysine demethylase Utx to promote 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

84 is widely expressed and required for placental and embryonic development (Sebastiano et al.,
85 2010; Wang et al., 2004). Circumventing the placental defects via tetraploid complementation
86 results in developmental arrest at E8.25 with no more than five somites (Sebastiano et al., 2010).
87 Oct1-deficient ESCs are morphologically normal and proliferate and self-renewal normally, but
88 upon differentiation show phenotypic and gene expression defects including decreased
89 developmental lineage-specific gene expression and elevated expression of developmentally
90 incorrect genes (Shen et al., 2017). Extended incubation of these cells supports an interpretation
91 that there is a failure to differentiate rather than a simple kinetic delay (Shen et al., 2017). The
92 underlying molecular mechanisms by which Oct1 regulates lineage differentiation are unknown.
93 Here, using MD differentiation of *Oct1*/*Pou2f1* inducible-conditional ESCs, we show that
94 pluripotent cells lacking Oct1 fail to form terminally differentiated myotubes. Differentiation is
95 restored by Oct1 retroviral complementation. Bulk RNA-seq profiling early in the differentiation
96 timecourse identifies gene expression abnormalities early in MD differentiation that predict the
97 later phenotypic defects. Single-cell RNA sequencing reveals that cells lacking Oct1 accumulate
98 abnormal cell populations associated with epithelial characteristics, oxidative stress and early
99 MD, while almost completely failing to reach somite-stage differentiation. Pseudotime analysis
100 reveal increased predilection to proceed down incorrect developmental trajectories, and "fuzzy"
101 differentiation programs. We show that Oct1 interacts with components of the Utx (Kdm3a)
102 H3K27me3 demethylase complex, and recruits Utx to developmentally appropriate target genes.
103 At tested targets this results in specific removal of H3K27me3 and de-repression. Binding sites
104 for Oct1 and Oct4 in pluripotent cells in which Oct1 binding is carried forward during MD
105 differentiation are also enriched in sites for Smad transcription factors, key mediators of MD
106 specification. Oct1 and Smad3 interact in differentiating cells, providing a means of restricting
107 Oct1's activating potential to MD-specific genes. The cumulative results support a model in which
108 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

141 To identify how Oct1 loss perturbs MD differentiation, we performed bulk RNA-seq using parental
142 and Oct1 cKO cells. Three replicates were performed per condition. ~1700, ~800 and 3,300
143 significantly differentially expressed genes were identified at D0, D3 and D6, respectively (FDR ≤
144 0.05; -1 < log2FC < 1, Supplementary Table 1). Euclidean distance analysis reveals tight
145 correlations between replicates and conditions at D0 and D3, but divergence at D6 relative to the
146 other conditions, and devergence between parental and cKO D6 replicates (Supplemental Figure
147 1). Unsupervised hierarchical clustering reveals groups of genes regulated similarly between
148 parental and cKO cells, and others with differential expression (Figure 1E). Cluster 2 for example
149 shows strong induction in differentiating parental but failed induction in cKO D6 cells. Example
150 genes in this set include *Pax3*, *Pax7* and *Myog* (Figure 1E, box). GO analysis of the differentially
151 expressed genes in cluster 2 reveals association with differentiation to neural tube
152 (neuroectoderm) and somite (Figure 1F). By contrast, clusters 1 and 4, which were associated
153 with aberrant gene expression in differentiating cKO cells, were associated with lineage-
154 inappropriate terms such as primordial germ cells, vascular system and stromal cells (Figure 1F
155 and not shown). To identify potential regulators of cluster 2 genes with failed induction in
156 differentiating cKO cells, we queried the ChIP Enrichment Analysis (ChEA) database (Lachmann
157 et al., 2010). Cluster 2 genes tend to be bound by the polycomb repressor compex 2 (PRC2)
158 components Suz12 and Mtf2 (Figure 1F). Example tracks are shown in Figure 1G for *Pou5f1*
159 (pluripotency), *Tbxt* (*T*/*Brachyury*, cluster 3) and *Pax7* (cluster 2). The retained D3 expression of
160 *Pou5f1*, which encodes Oct4, in both WT and cKO cells provides a likely explanation for the tight
161 correlations between gene expression states at D0 and D3 with divergence at D6 (Supplemental
162 Figure 1), as the presence of Oct4 may mask loss of Oct1 until D6. These data indicate that Oct1
163 loss results in defective induction of developmental genes that are also regulated by chromatin-
164 modifying activities that act on H3K27me3.
165
166 **Differentiating ESCs lacking Oct1 proceed down abnormal developmental trajectories**
167 To investigate perturbations in cell populations and gene expression at single-cell resolution, we
168 performed single-cell RNA-seq (scRNA-seq) using parental and cKO ESCs, and cells
169 differentiated towards MD for 3 or 6 days. Cell numbers ranged between 1000 and 2500, with
170 reads per cell ranging between 90,000 and 200,000 (see methods). Data from the different
171 conditions were subjected to graph-based clustering and uniform manifold approximation and
172 projection (UMAP) to visualize clusters of cells with similar gene expression. Integrated analysis
173 of undifferentiated parental ESCs and parental cells at the two differentiation timepoints identifies

174 a range of expected cell populations, from naïve and primed pluripotency to neuromesodermal,
175 paraxial MD and early somite (Figure 2A). These populations recapitulate developmental stages
176 associated with induction of markers associated with early MD (*Fgf17* and *Tbxt*/*T*/*Brachyury*;
177 cluster 2 and 7), neuromesoderm (*Hoxc9*, *Pax3*, *Sox2*), paraxial MD (*Meox1*, *Twist1*, *Pax3*) and
178 early somite and neuronal progenitors (*Cxcr4, Uncx*) (Figure 2B, Supplementary Figure 2 and not
179 shown). These results provide a single cell-resolution map of early MD differentiation.

180 Next, we compared the parental controls to cKO cells to identify changes in cell
181 populations and gene expression. D0 and D3 cKO cells show few differences from parental, Oct1
182 sufficient cells (Supplementary Table 2). In contrast, D6 cKO cells show poor differentiation
183 capacity with reductions in key mesodermal clusters (Figure 2C). For example, populations
184 characterized by neuromosodermal and paraxial MD gene expression represent the largest two
185 parental cell clusters, accounting for 14.7 and 12.9% of cells, respectively, while populations
186 associated with blood progenitors and somites account for 5.7% and 5.6% of cells (Figure 2C, left
187 panel). The complexity in the populations is consistent with findings that somites are derived from
188 multiple transcriptional trajectories including early paraxial mesoderm and neuromesodermal
189 progenitors (Guibentif et al., 2021). In contrast, cKO D6 cells show increases in clusters that retain
190 epithelial characteristics and dramatically decreased percentages of neuromesodermal
191 progenitors (2.3%, a six-fold decrease), paraxial MD (7.5%, two-fold), and both blood progenitors
192 (1.1%, >5-fold) and somites (0.5%, >10-fold, Figure 2C, right panel). This failure to produce more
193 differentiated MD lineage cells is consistent with findings that ESCs lacking Oct1 differentiate
194 poorly into myotubes (Figure 1A).

195 Gene expression comparisons between parental and cKO paraxial mesoderm clusters
196 reveals that cKO cells fail to robustly induce lineage-appropriate genes such as *Pax3* and *Pax7*,
197 and inappropriately upregulate lineage-inappropriate markers such as the epithelial-specific
198 genes *Krt8* and *Krt18* (Figure 2D). These results show that Oct1 is necessary for accurate
199 mesoderm differentiation and to suppress expression of genes for alternative lineages.

200 Pseudotime analysis of scRNA-seq data allows multiple differentiation timepoints to be
201 overlaid with defined starting (ESC) and endpoint (somite) populations. Parental control cells
202 progress through a largely linear pathway with one minor branch retaining an inappropriate
203 epithelial fate (Figure 3A, top panel). In contrast, differentiating cKO ESCs show a larger
204 proportion inappropriate branching into alternative developmental trajectories (bottom panel),
205 consistent with the diminished developmental progression of paraxial MD to the somite stage,
206 and consistent with enrichment of cells that inappropriately maintain an epithelial state (Figure 2C

207 and Supplementary Figure 2). We also examined pseudotemporal gene expression using specific
208 genes associated with pluripotency and MD development. In this analysis, the position of each
209 cell in pseudotime is shown on the X-axis and the degree of expression for the indicated gene on
210 the Y-axis (Figure 3B). Parental cells show robust early expression of genes associated with
211 pluripotency such *as Klf4* that lose expression roughly mid-way through pseudotime. *Tbxt* is
212 transiently induced, while the early somite markers *Pax7* and *Cxcr4* are efficiently and
213 coordinately induced later (Figure 3B, top panel). In contrast, cKO cells exhibit inappropriately
214 prolonged expression of *Klf4* and *Tbxt*, and largely fail to upregulate *Pax7* and *Cxcr4* (bottom
215 panel). We used pseudotemporal ordering to visualize expression of the 2000 most variable
216 (dynamic) genes in parental cells, and compared expression of the same genes in the same order
217 in cKO cells (Figure 3C). This analysis revealed clusters associated with pluripotency (*Klf4*,
218 *Nanog, Pou5f1*), early MD differentiation (*T*), paraxial mesoderm (*Rspo3, Meox1*), and early
219 somites (*Dll1, Pax3*). The pseudotime gene expression pattern for cKO cells revealed largely
220 normal expression of early gene expression markers, but loss of temporal coherence at
221 subsequent steps. These "fuzzy" patterns of gene expression begin early in differentiation, with
222 for example prolonged expression of genes such as *Tbxt* and noncoherent expression of genes
223 associated with paraxial mesoderm such as *Meox1* and *Rspo3*. The diffuse pattern is different
224 from what would be expected with a kinetic delay, which would cause a rightward shift. In addition,
225 the induction of somitic genes normally expressed later in development such as *Dll1*, *Pax3* and
226 *Pax7* was poor in cKO cells (Figure 3C).

227 RNA velocity analysis allows developmental directionality to be inferred relative to other
228 cells by comparing the ratio of unspliced, newly-synthesized pre-mRNA to mature spliced mRNA.
229 A vector is then assigned to each cell that indicates developmental direction and magnitude
230 relative to the other cells. We clustered the cells separately by genotype, resulting in different
231 patterns between genotype and compared to cells clustered together (Figure 3D). Applying
232 velocity to this clustering, we found that D6 differentiated parental cells form discrete clusters with
233 dynamic developmental progression, e.g. paraxial MD and somite (Figure 3D, marked by long
234 arrows). cKO cells by contrast are marked by stationary profiles indicative of failed differentiation
235 potential, with cells progressing primarily towards a poorly differentiated state characterized by
236 an aberrant epithelial state and multiple lineage markers (Figure 3D, epithelial clusters).
237 Cumulatively, the data demonstrate that Oct1 is required for efficient early adoption of the MD
238 differentiation program, with Oct1-deficient cells unable to canalize appropriate lineage
239 specification programs.

240

241 **Oct1 occupies developmental genes in ESCs**

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

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

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

273 differentiation (Figure 4A). In contrast, 807 shared Oct4/Oct1 peaks in pluripotent cells continue
274 to be bound by Oct1 at D6. Analysis of these peaks using GREAT (McLean et al., 2010) identifies
275 enrichment for oxidative stress, ribosomal and mitochondrial organization, Notch signaling and
276 post-implantation development including somite formation (Supplementary Figure 4A).
277 Additionally, >6000 peaks become uniquely bound by Oct1 at MD D6 (Figure 4A). >2300 peaks
278 are uniquely bound at D3 (Supplementary Figure 4B). We compared Oct1 target gene binding at
279 D6 in parental cells to genes that were significantly down-regulated in cKO cells at the same
280 timepoint (FDR < 0.05; negative log2(FC)), observing significant overlap (Figure 4B).

281 To pattern Oct1 and Oct4 occupancy during MD differentiation, we applied hierarchical
282 clustering. We used a union of all peaks, as opposed to peak intersect in Figure 4A, to assess
283 varying levels of Oct1 enrichment over the course of differentiation. Three major clusters of
284 occupied targets were identified, two of which show decreased or static Oct1 binding over time
285 (Figure 4C, clusters 1 and 2). Gene ontogeny (GO) analysis indicates involvement in signaling,
286 mitochondrial function, stem cell differentiation and regulation of Notch and TGFβ signaling
287 (Figure 4C). Oct1 and Oct4 enrichment at a cluster 1 gene encoding a pluripotency factor (*Pou5f1*)
288 and two silent, developmentally poised cluster 2 genes encoding MD/myogenic transcription
289 factors (*Pax3* and *Myog*) is shown in Figure 4D. These latter genes are frequently associated with
290 bivalency. To identify bivalent genes, we intersected ESC H3K27me3- and H3K4me3-enriched
291 ChIP-seq peaks from the ENCODE database. This procedure identifies 3861 bivalent genes
292 (Supplementary Table 4). To query Oct1 occupancy on these genes during differentiation, we
293 intersected the bivalent gene dataset with Oct1 ChIP-seq in pluripotent and MD-differentiated
294 parental cells, observing an increase in binding over time (Figure 4E, green and blue lines). A
295 similar analysis at D6 using just MD-specific bivalent genes (generated by intersecting the bivalent
296 gene dataset with MD development GO:0007498, Supplementary Table 5) shows even stronger
297 Oct1 enrichment (Figure 4D, purple line). These findings indicate that Oct1 robustly binds to
298 bivalent genes both in pluripotent cells and their differentiated progeny, with binding increasing at
299 lineage-specific genes during differentiation.

300 In contrast, Oct1 binding in Figure 4C increases with differentiation in cluster 3. GO
301 analysis indicates that genes near these peaks are associated with chromatin-modifying activities
302 including histone acetyltransferase and polycomb group complexes. Oct1 and Oct4 enrichment
303 at example cluster 3 genes encoding polycomb complex members (*Ezh2*, *Suz12*, *Ring1* and
304 *Ezh1*) is shown in Supplementary Figure 5A. These genes show a gain in Oct1 occupancy as
305 cells progress from pluripotency through D3 to D6 of MD differentiation. In the scRNA-seq UMAP

306 projection, *Ezh2* expression is downregulated in cKO cells at D6, in particular in paraxial
307 mesoderm (Supplementary Figure 5B-D). These data indicate that during MD differentiation, Oct1
308 directly binds to, and mediates the induction of, genes encoding epigenetic regulators associated
309 with H3K27me3.
310
311 **Oct1 Recruits Utx to the *Pax3* Target Gene to Remove H3K27me3.** Lineage-appropriate
312 poised developmental genes become activated in part via the removal of repressive H3K27me3
313 marks (Bernstein et al., 2006; Dhar et al., 2016a). Our findings indicate that genes with failed
314 induction in differentiating cKO cells tend to be bound by chromatin-modifying complexes that act
315 on H3K27me3 (Figure 1F). We hypothesized that Oct1 locally recruits H3K27-specific
316 demethylases to lineage-specific genes to mediate their induction during differentiation. One such
317 demethylase is Utx/Kdm6a (Cloos et al., 2008), which has been shown to regulate MD
318 differentiation (Wang et al., 2012). To test the association between Oct1 binding and Utx, we also
319 performed ChIP-seq using H3K27me3 and Utx antibodies at D6 of differentiation of parental, Oct1
320 sufficient cells. ~12,000 H3K27me3 peaks and ~12,000 Utx peaks were identified, corresponding
321 to ~11,000 genes within 20 kb (Supplementary Table 3). Unsupervised hierarchical clustering
322 together with Oct1 peaks from the same cells identifies shared and distinct peaks, including weak
323 Oct1 binding events with strong Utx and H3K27me3 association (Figure 5A, cluster 3), and strong
324 Oct1 binding events associated with weaker Utx and lack of H3K27me3 (cluster 4). This latter
325 cluster includes genes induced in MD such as *Pax3, Pax7* and *Myog*. We interpret cluster 3 to be
326 developmental genes that remain poised at this timepoint, and cluster 4 to be lineage-appropritate
327 genes that have been activated. GO terms associated with these clusters were enriched for
328 development and musculoskeletal abnormalities (Figure 5B). Intersecting the Utx and Oct1 ChIP-
329 seq peaks identifies a high degree of overlap, with ~70% of Oct1-bound peaks also associating
330 with Utx (Figure 5C).
331 To test if cKO cells inappropriately retain H3K27me3 at lineage-appropriate genes during
332 MD differentiation, we performed ChIP-qPCR using primers flanking a strong Oct1 peak on the
333 *Pax3* promoter. *Pax3* is induced by D6 of MD differentiation, and reduced in Oct1 cKO cells at
334 this timepoint (Figure 1E and Figure 3C). As expected, H3K27me3 was robust and equivalent in
335 parental and cKO D0 undifferentiated cells. D6 parental cells showed reduced H3K27me3, while
336 in contrast, cKO cells inappropriately retained elevated H3K27me3 (Figure 5D). This failure to
337 remove H3K27me3 resulted in ~3-fold higher H3K27me3 enrichment in differentiated cKO relative
338 to parental cells. Global H3K27me3 levels are unchanged at the same timepoint, as evidenced

339 by immunofluorescence (Figure 5E). Next, we performed Utx ChIP-qPCR with the same primers.
340 Utx is enriched at the *Pax3* promoter at D6 of MD differentiation, validating the ChIP-seq
341 enrichment in parental cells. Further, Utx showed reduced enrichment in cKO cells (Figure 5F).
342 These results indicate association of Utx and Oct1 at *Pax3*, and specific defects in removing
343 H3K27me3 from *Pax3* in differentiating Oct1-deficient cells.
344 We then determined if Oct1 and Utx interact using extracts from undifferentiated and D6-
345 differentiated parental ESCs and co-immunoprecipitation. The Oct1-Utx interaction is present
346 even in ESCs, and is maintained during differentiation (Figure 5G). These results support the
347 interpretation that Oct1 interacts with Utx and recruits it to target genes during differentiation to
348 help mediate their activation. Utx recruitment by Oct1 potentially explains the failed induction of
349 lineage-appropriate genes such as *Pax3*, and the failed removal of H3K27me3 in cKO cells
350 (Figure 5D). Oct1, Oct4 and Utx binding to *Pax3* is shown in Figure 5H.
351
352 **Oct1 interacts with Smad3, and cooperates with Smad proteins at *Myog* in cells**
353 **differentiating towards MD.** The broad expression of Oct1 raised the question of how Oct1
354 mediates gene activation specifically at MD lineage-specific. Chromatin accessibility, lineage-
355 specific signaling and co-bound transcription regulators may provide this specificity. We
356 performed motif analysis (Heinz et al., 2010) using DNA sequences ±100 bp from the center of
357 the 807 peaks co-bound by Oct1 and Oct4 in pluripotent cells that remain Oct1-bound during MD
358 differentiation (Figure 4A). This procedure identifies not only Oct1 sites, but also binding sites for
359 Smads, the terminal transcription regulators of TGFβ/Nodal/BMP signaling (Figure 6A). TGFβ
360 signals and downstream Smad transcription factor activity are critical for MD induction (Conlon et
361 al., 1994; Zhou et al., 1993). A study identified Oct1 transcription factor motifs near binding events
362 for zebrafish Smad2/3 and further showed that ectopic expression of the zebrafish Oct1 ortholog
363 enhances mesoderm induction, that zebrafish Smad2 and Oct1 physically interact, and that the
364 two factors cooperate to enhance transcription (Liu et al., 2011). Mammalian Smad3 also interacts
365 with Oct4 and co-occupies target genes in pluripotent cells (Mullen et al., 2011). Consistent with
366 these findings, Oct1 antibodies efficiently co-precipitate Smad3 in D6 MD-differentiated cells
367 (Figure 6B). *Smad3* is expressed in undifferentiated parental populations, but further induced by
368 MD differentiation (Figure 6C).
369 To gain more insight into mechanism, we tested if Oct1 and Smad proteins synergize to
370 regulate gene expression. The *Myog* gene fails to be induced in D6 differentiating Oct1 cKO ESCs

371 and contains a strong Oct1 binding event overlapping with an annotated distal enhancer ~3 kb
372 upstream of the *Myog* TSS (Figure 6D). The region is also located within a known super-enhancer
373 associated with myotubes (Hnisz et al., 2013). We cloned an 85 bp region of this enhancer
374 centered on the Oct1 ChIPseq peak together with the CMV core promoter upstream of a
375 nanoLuciferase reporter vector (Figure 6E). This region contains two octamer sites and two Smad
376 sites, and overlaps with an ENCODE distal enhancer signature (Abascal et al., 2020). The
377 reporter vector was co-transfected with a plasmid constitutively expressing secreted mCherry
378 (Wider and Picard, 2017) as an internal standard into MEFs lacking Oct1 (Wang et al., 2004)
379 (Figure 6D). Cells were also co-transfected with a construct encoding mouse Oct1 and/or treated
380 with recombinant TGFβ1 to supply Oct1 and SMAD activity. The two treatments together generate
381 cooperative luciferase activity (Figure 6F). For example, constructs containing only active Smad
382 or Oct1 binding sites (Figure 6E , green and blue) displayed little more activity than constructs
383 with no sites (purple), but the two sites together generate significantly increased activity (red). The
384 affect was attenuated in the absence of supplemented Oct1 or Smad activation. These results
385 indicate that Oct and Smad sites at the *Myog* enhancer cooperate to drive gene expression.
386

387 **Discussion**

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

407 The central role of Oct1 in "canalizing" differentiating pluripotent cells in early steps of
408 mesodermal specification was shown using single-cell RNA-seq with Oct1 cKO ESCs. In the
409 absence of Oct1, cells undergoing differentiation retain an epithelial state, and achieve
410 neuromesodermal and somite-stage gene expression patterns poorly and in reduced numbers.
411 Increased populations of cells appear associated with oxidative stress, consistent with findings
412 that cells lacking Oct1 have elevated reactive oxygen species levels (Tantin et al., 2005). cKO
413 cells mis-express developmentally inappropriate genes and undergo inappropriate developmental
414 branching towards poorly differentiated states marked by epithelial gene expression and oxidative
415 stress. Unlike the expected effect of loss of lineage-specific master regulators where cell fate
416 often shifts to a specific alternate lineage, these cells lack clear lineage signatures. The induction
417 of genes important early in MD differentiation such as *Tbxt* becomes weaker and loses temporal
418 coherence. Later, there is failed induction of genes such as *Pax7* and *Cxcr4*.

419 In MD-differentiating cells, Oct1 occupies a subset of genes bound by Oct4 in pluripotent
420 cells. These include lineage-specific developmental mediators and genes encoding chromatin-
421 modifying enzymes. Oct1 occupancy on these genes increases with differentiation, suggesting a
422 critical role in canalization of cell fate. We find that Oct1 recruits Utx to lineage-specific, Oct1-
423 bound targets such as *Pax3*. The lack of Utx recruitment to lineage-specific genes in cKO cells is
424 consistent with the abnormal retention of H3K27me3 at the promoter and the failed transcriptional
425 upregulation. No global changes in H3K27me3 were observed, consistent with recruitment by
426 Oct1 to mediate focal removal (Wang et al., 2012). An "anti-repression" mechanism for Oct1
427 involving the removal of repressive chromatin marks has been described for Oct1 before, for
428 H3K9me2 (Shakya et al., 2011). To our knowledge this is the first description of a role for Oct1
429 and K3K27me3. Loss of Utx allows for maintainance of an undifferentiated state, but interferes
430 with MD differentiation (Dhar et al., 2016b).
431 One mechanism that allows Oct1 to gain specificity for MD-appropriate targets is
432 collaboration with Smad3. Smad transcription factor binding sites are enriched near sites of Oct1
433 binding during MD differentiation. TGFβ signals drive Smad transcription factor binding and are
434 critical for MD specification (Liu et al., 2011; Mullen et al., 2011). Co-IP experiments in D6 MD-
435 differentiated cells show an interaction between Oct1 and Smad3, consistent with prior findings
436 in zebrafish (Liu et al., 2011). A model for Oct1's function at lineage-specific genes during MD
437 specification and later differentiation is shown in Figure 7. In this model, Oct1 and Smad form
438 cooperative complexes at MD-specific genes, and Utx recruitment to Oct1 allows for loss of the
439 repressive H3K27me3 mark and gene activation. Subsequently, other transcription factors (e.g.,
440 MyoD) act as primary "on" switches to induce gene expression. Oct1 also binds and induces the
441 expression of genes encoding PRC complex members such as *Ezh2* in MD-specific clusters in
442 parental but not cKO cells. The increased expression of PRC components may solidify lineage
443 specification by aiding the repression of the large number of genes specific to alternative lineages
444 (Collinson et al., 2016).
445 Under MD differentiation conditions, exogenous Oct1 increases expression of the terminal
446 differentiation markers *Myod* and *Myog*, while decreasing the early lineage-specification marker
447 *Pax3*, which is transiently expressed during MD differentiation and is necessary for later
448 expression of myogenic genes. Because *Pax3* is no longer expressed in terminally differentiating
449 cells, these results suggest that ectopic Oct1 may enable transit through a *Pax3*-expressing
450 intermediate to potentiate productive terminal differentiation. More investigation into this pathway

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

453 **Figure Legends**

454

455 **Figure 1. Failed MD/muscle lineage induction in differentiating Oct1 cKO ESCs.** (**A**)
456 Embryonic myosin heavy chain (MyHC-emb) expression alone and merged with DAPI is shown
457 using parental (top panel) or cKO (bottom panel) cells at MD differentiation D11. (**B**) Relative
458 mRNA expression levels of the myogenic genes *Myod1* and *Myog* in parental and cKO ESCs
459 differentiated for 19D. Data represent an average of biological triplicates. Error bars depict ±SEM.
460 (**C**) RT-qPCR for the myogenic genes *Myod1*, *Myog* and *Pax3* in parental and cKO cells
461 transduced with retroviruses encoding Oct1 or empty vector. Transduced cells were selected
462 using puromycin for 48 hr prior to differentiation for 11D. Empty vector values were set to 1. Data
463 represent an average of biological triplicates. Error bars depict ±SEM. (**D**) Immunoblot showing
464 ectopic Oct1 expression in parental and cKO cells. β-actin is shown as a loading standard. (**E**)
465 Bulk RNA-seq heatmap of differentially expressed genes (parental vs. cKO) at D0, D3 and D6 of
466 MD differentiation is shown. Cluster 2 shows poor gene induction in the Oct1-deficient condition.
467 Representative cluster 2 genes are shown at right. (**F**) Jensen Tissue and ChIP-X Enrichment
468 Analysis (ChEA) query results for Cluster 2 in (B) are shown. (**G**) Example RNA-seq genome
469 tracks (*Pou5f1*, *Tbxt*, *Pax7*) are shown. *Pax7* is an example cluster 2 gene. Y-axes were group-
470 autoscaled for each gene.

471

472 **Figure 2. Differentiating Oct1-deficient ESCs canalize poorly into mesodermal lineages.** (**A**)
473 UMAP projection of scRNA-seq data from superimposed parental (Oct1 sufficient)
474 undifferentiated ESCs, and parental cells early during MD differentiation (days 3 and 6). Clusters
475 of cells were labeled based the expression of developmental markers as in (B). MD-associated
476 clusters are shown in bold. Three combined replicate plates were used for the analysis. (**B**) Violin
477 plots showing gene expression levels of key developmental markers by cluster. Data were log-
478 normalized for each cell using the natural logarithm, scaled and averaged using mean(expm1(x)).
479 (**C**) Comparative UMAP projections of integrated D6 parental and Oct1-deficient (cKO) scRNA-
480 seq populations. Clusters were labeled computationally and identified based on gene expression.
481 Relative frequencies are shown. MD lineage-associated clusters are shown in bold. (**D**)
482 Differential gene expression analysis of the neuromesodermal cluster shown as a scatter plot.
483 Red dots depict significantly differentially expressed genes based on FDR corrected $p < 0.05$ and
484 fold change >1.2. Example differentially expressed genes are labeled.

485

486 **Figure 3. Oct1-deficient cells show perturbed developmental trajectories.** (**A**) Pseudotime
487 analysis of pluripotent and differentiating parental (top panel) and cKO (bottom panel) cells.
488 Colors correspond to the time point at which cells were collected (red: D0, green: D3, blue: D6).
489 (**B**) *Klf4*, *Tbxt*, *Pax7* and *Cxrc4* mRNA expression across pseudotime in parental (top panels) and
490 cKO (bottom panels). Black trendline represents an average expression for a given gene across
491 all populations. (**C**) Heatmap depicting expression of the 2000 most dynamically expressed genes
492 (based on FindVariableFeatures function, Seurat) in parental D6 cells. Gene expression was
493 plotted as a heat map across pseudotime in parental (left panel) and cKO (right panel). Dynamic
494 genes were first hierarchically clustered in parental cells to cluster groups of genes that behave
495 similarly in pseudotime, then plotted in the same order in cKO cells. (**D**) Velocity gene expression
496 analysis of parental (left panel) and cKO (right panel) cells at differentiation day 6. Arrows point
497 toward cells with gene expression closest to the future state of each cell based on
498 unspliced/spliced transcripts. Arrow length represents magnitude. Clusters in bold are associated
499 with MD differentiation.

500

501 **Figure 4. Oct1 co-occupies target sites with Oct4 in ESCs, and regulates their expression**
502 **during differentiation.** (**A**) Venn diagram showing common and unique Oct4 and Oct1 binding
503 sites based on ChIP-seq in parental ESCs and at D3 and D6 of mesodermal differentiation. (**B**)
504 Comparison of Oct1 binding events and genes down-regulated by Oct1 loss at MD differentiation
505 day 6. Target genes (< 20 kb) were used rather than peaks, explaining the differing numbers
506 compared to panel (A). (**C**) A matrix of Oct4 and Oct1 ChIP-seq enrichment 2 kb around peak
507 centers was computed for the merged peak list and shown as heatmap. Positions in color show
508 high enrichment and white shows no enrichment. Associated GO terms are shown at right. (**D**)
509 Example cluster 1 and 2 genome tracks: *Pou5f1* (cluster 1), *Pax3* (cluster 2) and *Myog* (cluster
510 2). (**E**) Oct1 enrichment based on tag density at peak center at annotated bivalent genes in ESCs
511 (red), and at MD differentiation D3 and D6 (green and blue). An additional analysis was performed
512 for MD-specific bivalent genes at MD differentiation D6 (purple).

513

514 **Figure 5. Oct1 recruits Utx to demethylate H3K27me3 at lineage-appropriate genes.** (**A**)
515 Oct1, H3K27me3 and Utx ChIP-seq enrichment 2 kb around peak centers was computed for the
516 merged peak list and shown as heatmap. (**B**) Human Phenotype Ontology terms for Cluster 3 and
517 4 genes from (A) are shown. (**C**) The Oct1 and Utx peak lists were intersected (overlap ≥ 1 bp)

518 and plotted as a Venn diagram. 69% of Oct1 peaks overlapped with Utx binding events. (**D**)
519 H3K27me3 ChIP-qPCR enrichment at the *Pax3* promoter in parental and cKO cells at MD
520 differentiation D0 (ESC) and D6. Normalized fold-enrichment is shown relative to both an isotype
521 control antibody and to a nonspecific genomic region encoding the 28S ribosomal subunit. Data
522 represent an average of 3 biological replicates. Error bars depict ±SEM. (**E**) H3K27me3
523 immunofluorescence images from D6 MD-differentiated parental and cKO cultures. Images were
524 collected a 40× magnification. (**F**) Utx enrichment at the *Pax3* promoter in parental and cKO cells
525 on differentiation D6. Fold-enrichment relative to an isotype control antibody and a nonspecific
526 28S region is shown. Data represent an average of 3 independent biological replicates. Error bars
527 depict ±SEM. (**G**) Utx immunoprecipitation followed by Oct1 immunoblot using parental ESCs, or
528 ESCs MD-differentiated for 6D. (**H**) Signal tracks (*Mm10* v.D191020) showing Oct4, Oct1 and Utx
529 enrichment at the *Pax3* locus 5' region. Y-axes were group-autoscaled for each gene. Positions
530 of identified HOMER motifs are shown below.
531
532 **Figure 6. Oct1 and Smad3 cooperate to drive expression of mesoderm-specific genes.** (**A**)
533 HOMER motif analysis of Oct1 peaks that are shared with Oct4 in ESCs, and maintained after D6
534 of MD differentiation. (**B**) Smad3 immunoblot using cell lysates immunoprecipitated with Oct1
535 antibodies, or rabbit IgG controls. D6 MD-differentiated cells were used. 20% input is shown (lane
536 1). (**C**) *Smad3* gene expression is shown in violin plots for parental ESCs at D6 of MD
537 differentiation. (**D**) Signal tracks (*Mm10* v. D191020) showing Oct1ChIP-seq enrichment at the
538 *Myog* locus. Shown above are RNA-seq tracks in differentiated and undifferentiated parental and
539 cKO cells. ENCODE-annotated regulatory elements are shown below. Pink: proximal promoter.
540 Yellow: distal enhancer. (**E**) Scheme for Oct1/Smad reporter assay. A segment of a mouse *Myog*
541 enhancer element containing multiple Oct and Smad motifs was cloned with the core CMV
542 promoter upstream of secreted nLuc and co-transfected into Oct1-deficient MEFs together with a
543 construct encoding constitutive secreted mCherry as a normalization control. Added TGFb1 and
544 co-transfected mouse Oct1 supply Oct1 and Smad3 activity. (**F**) Transfected WT or Oct/Smad
545 mutant *Myog* enhancer constructs were supplied with Oct1, recombinant purified TGFb1
546 treatment, or both. Secreted luciferase activity was assessed relative to secreted mCherry
547 expressed from a co-transfected plasmid. An average of experimental triplicates is shown. Error
548 bars denote ±SEM. For the situation in which both Oct1 and TGFb1 are both supplied, fold
549 changes relative to a double-mutant construct are also shown.

550

551 **Figure 7. Sequential model of bivalency resolution for lineage-appropriate (MD-specific)**
552 **genes.** Pluripotent cells co-express Oct1 and Oct4, which co-bind to poised targets. Upon loss of
553 pluripotency and Oct4, Oct1 continues to occupy these genes. TGFβ signals allow co-binding of
554 Oct1 and Smad proteins to MD-specific targets, recruitment of Utx and demethylation of
555 H3K27me3. Later, other transcription factors serve as primary "on" switches for muscle-specific
556 gene expression.

557 **Methods**

558 **High-throughput datasets.** The datasets generated during this study are available through the
559 GEO website [Series record GSE160943]. The record is in private status. The following secure
560 token has been created to allow reviewer access while it remains in private status:
561 ylqjokymttavpah. Should the manuscript be accepted for publication, we will direct GEO to release
562 the data publicly. ENCODE data sets for H3K27me3 and H3K4me3 in ESC-Bruce4 embryonic
563 stem cells are deposited under GSE31039.

564

565 **Sample size estimation and replicates.** No explicit power analysis was used. Sample sizes
566 were estimated based on calculated differences from pilot experiments. In general, a biological
567 replicate used independently cultured cells, whereas a technical replicate relied on a later aliquot
568 from a derived component, e.g. aliquoted RNA for qRT-PCR. No outliers were removed.

569

570 **Cell culture.** ESCs were cultured as previously described (Shakya et al., 2015) with 2i conditions:
571 ERK inhibitor PD0325901 (1 μM, LC Laboratories) and GSK3 inhibitor CHIR99021 (3 μM, LC
572 Laboratories). Cultures were maintained on irradiated feeders (ThermoFisher). Prior to all
573 experiments ESCs were plated on gelatin to deplete the feeders. For Oct1 deletion, cells were
574 treated with 4-hydroxytamoxifen and a completely yellow sub-colony was picked and expanded
575 as published (Shen et al., 2017). For MD differentiation, ESCs were plated on gelatin and cultured
576 as previously described (Chal et al., 2015). Briefly, parental and cKO cells were cultured in N2B27
577 medium supplemented with recombinant Bmp4 (Peprotech) for 2 d. After 48 hr, media was
578 changed to RDL (Rspo3, DMSO, LDN) medium. Cells were harvested 24 hr (day 3) or 96 hr (day
579 6) later. For muscle differentiation, cells were switched to HIFL (Hgf, Igf, Fgf, Ldn) medium and
580 cultured for 48 hr (day 8) after which medium was switched to 2% horse serum (ThermoFisher).
581 Cells were harvested on day 11 (overexpression experiments) or 19 (RT-qPCR).

582

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

587

588 **Retroviral Oct1 overexpression.** Oct1 was ectopically expressed in ESC cells using a previously
589 described retroviral vector (pBabePuro-Oct1) (Kang et al., 2009b). pBabePuro was used as an

590 empty vector control. The vector was co-transfected together with the pCL-Eco packaging plasmid
591 into HEK293T cells to generate retroviral particles. Retroviral supernatants were collected 48 hr
592 later, filtered through 0.45 μm filters and applied to ESCs cultures maintained on puromycin-
593 resistant feeder fibroblasts (ThermoFisher). The mixed population of cells was subjected to
594 selection with puromycin for 48 hr.
595
596 **RT-qPCR.** RNA was isolated using TRIzol (ThermoFisher). cDNA was synthesized using a
597 SuperScript Vilo cDNA Synthesis Kit (ThermoFisher). RT-qPCR oligonucleotide primers are listed
598 in Supplementary Table 6 and were confirmed to generate a single product of the correct size. To
599 ensure specific PCR amplification, every RT-qPCR run was followed by a dissociation phase
600 analysis (denaturation curve) to confirm the presence of single amplified peaks.
601
602 **Bulk RNA-seq.** RNA was prepared from three independent cultures of undifferentiated or 3 d and
603 6 d MD-differentiated parental or cKO ESCs. Poly(A) RNA was purified from total RNA samples
604 (100–500 ng) with oligo(dT) magnetic beads, and stranded mRNA sequencing libraries were
605 prepared as described using the Illumina TruSeq mRNA library preparation kit and RNA UD
606 Indexes. Molarity of adapter-modified molecules was defined by qPCR using the Kapa
607 Biosystems Library Quant Kit. Individual libraries were normalized to 1.3 nM. Sequencing libraries
608 were chemically denatured and applied to an Illumina NovaSeq flow cell using the NovaSeq XP
609 chemistry workflow. Following transfer of the flowcell to an Illumina NovaSeq instrument, 2×51
610 cycle paired-end sequence was performed using a NovaSeq S1 reagent kit. Between 13 and 18
611 million paired-end reads were generated for each condition. More than 99% of aligned reads
612 mapping to the correct strand.
613
614 **Bulk RNA-seq analysis.** The Illumina adapter sequence was trimmed using cutadapt version
615 1.16. Fastq data quality were checked using Fastqc verision 0.11.5. Reads were aligned to the
616 mouse *Mm10* genome using STAR version 2.7.3a in two-pass mode. Aligned reads were checked
617 for quality using the Picard tools' CollectRnaSeqMetrics command to count the number of read-
618 matching exons, UTRs, introns and intergenic regions, and to calculate normalized gene
619 coverage from the top 1000 expressed transcripts. Between 13 and 18 million paired-end reads
620 were generated for each condition, with >99% of aligned reads mapping to the correct strand.
621 Differentially expressed genes were identified using a 5% FDR with DESeq2 version 1.24.0 (Love

622 et al., 2014). Genes with a count of at least 50 in one or more samples were tested. Genes
623 showing at least 2.5-fold change of expression at an adjusted $p<0.01$ were selected as
624 differentially expressed. Figures were generated in R version 4.0.0 using functions from ggplots
625 libraries and pheatmap.
626
627 **Single-cell RNA-seq.** Single cell transcriptomes were analyzed as described previously
628 (Dell'Orso et al., 2019). The 10X Genomics Chromium Single Cell Gene Expression Solution with
629 3' chemistry, version 3 (PN-1000075) was used to tag individual cells with 16 bp barcodes and
630 specific transcripts with 10 bp Unique Molecular Identifiers (UMIs) according to manufacturer
631 instructions. Briefly, single-cell suspensions were isolated using trypsinization and resuspension
632 in PBS with 0.04% BSA (ThermoFisher). Suspensions were filtered through 40 μm cell strainers.
633 Viability and cell count were assessed using a Countess I (ThermoFisher). Equilibrium to targeted
634 cell recovery of 6,000 cells along with Gel Beads and reverse transcription reagents were loaded
635 to Chromium Single Cell A to form Gel-bead-in Emulsions (GEMs). Within individual GEMs, cDNA
636 generated from captured and barcoded mRNA was synthesized by reverse transcription at 53°C
637 for 45 min. Samples were then heated to 85°C for 5 min. Subsequent A tailing, end repair, adaptor
638 ligation and sample indexing were performed in bulk according to manufacturer instructions. The
639 resulting barcoding libraries were qualified on Agilent Technology 2200 TapeStation system and
640 subjected to qPCR using a KAPA Biosystems Library Quantification Kit for Illumina Platforms
641 (KK4842). The multiple libraries were then normalized and sequenced on an Illumina NovaSeq
642 6000 using the 2×150 PE mode.
643
644 **Single-cell RNA-seq data processing and clustering.** Sequences from the Chromium platform
645 were de-multiplexed and aligned using CellRanger ver. 3.1.0 (10X Genomics) with default
646 parameters mm10-3.0.0. Clustering, filtering, variable gene selection and dimensionality
647 reduction were performed using Seurat ver.3.1.5 (Stuart et al., 2019) according to the following
648 workflow: 1, Cells with <300 detected genes and >10000 genes were excluded further analysis.
649 2, Cells with <12% UMIs mapping to mitochondrial genes were retained for downstream analysis.
650 3, The UMI counts per ten thousand were log-normalized for each cell using the natural logarithm.
651 4, Variable genes (2000 features) were selected using the FindVariableFeatures function. 5,
652 Common anchors between the three parental timepoints (Figure 2A) or parental and cKO D6
653 (Figure 2C) were identified using FindIntegrationAnchors function that were further used to

654 integrate these sets. 6, Gene expression levels in the integrated set were scaled along each gene
655 and linear dimensional reduction was performed. The number of principal components was
656 decided through the assessment of statistical plots (JackStrawPlot and ElbowPlot). 7, Cells were
657 clustered using a by a shared nearest neighbor (SNN) modularity optimization-based clustering
658 algorithm and visualized using two-dimensional uniform manifold approximation and projection
659 (UMAP). 8, Cluster identities were defined based on the distribution of the specific markers.
660 Differentiational gene expression analysis between the parental and cKO clusters was performed
661 using FindMarkers. Genes with adjusted $p<0.01$ were marked red on scatter plots. Using this
662 analysis, the cell numbers were between 1000 and 2500. Reads/cell ranged between 90,000 and
663 200,000. Genes/cell ranged between 2,300 and 7,000. The percentage of reads mapping to
664 reference mouse genome was 90% or higher, and the reads confidently mapping to exonic
665 regions was 70% or higher.
666
667 **Pseudotime and Velocity analysis.** Trajectory analysis of scRNA-seq was performed using
668 Monocle v.2.16.0 (Trapnell et al., 2014). Parental and cKO sets were filtered using the same
669 parameters as above and merged to generate WT and cKO sets. Cells were ordered based on
670 gene lists for the ESC (beginning) and somite (end) clusters in parental UMAP (Figure 2A). Next,
671 we performed dimensional reduction using the DDRTree method to visualize the dataset, ordered
672 the cells by global gene expression levels, and visualized the trajectory of the dataset. Veolocity
673 analysis was performed using the velocyto package (La Manno et al., 2018). Loom files were
674 produced using following paramters: velocyto run10x -m mm10. rmsk.gtf genes.gtf. Gtf files were
675 produced from the Cell Ranger pipeline. Velocity embeddings were produced using the velocyto.r
676 and SeuratWrappers packages. Matrices were filtered using following parameters:
677 nFeature_spliced > 300, nFeature_spliced < 10000, nFeature_unspliced > 200,
678 nFeature_unspliced < 6000, percent.mt < 12. Velocity was calculated using RunVelocity using
679 following paremeters: deltaT = 1, kCells = 25, fit.quantile = 0.02. Velocity embedding were
680 projected on T-SNE maps using the show.velocity.on.embedding.cor function.
681
682 **ChIP.** ChIP-qPCR and ChIP-seq were performed as previously described (Perovanovic et al.,
683 2016). Briefly, WT and cKO cells were crosslinked with 1% formaldehyde for 10 min and
684 quenched for 5 min using 2.5M glycine. Culture plates were washed using ice cold PBS and cells
685 were harvested by cell scaping. Cells were lysed in Farnham buffer (5 mM Pipes pH 8.0, 85 mM
686 KCl, 0.5% NP-40) and subsequently in RIPA buffer (phosphate-buffered saline, 1% NP-40, 0.5%

687 sodium deoxycholate, 0.1% SDS). Chromatin was sheared using a Covaris sonicator for 5 min
688 (30 sec on/30 sec off) with amplitude=40. Correct chromatin fragmentation was confirmed using
689 1% agarose gels. 50 µg of chromatin was subjected to IP overnight at 4°C with 4 µg of anti-Oct1
690 (Novus Biological), Oct4 (Santa Cruz) or H3K27me3 (Milipore) antibodies. As a control, we used
691 5 µg of sheared, non-precipitated input chromatin. Samples were incubated with protein G
692 magnetic beads (ThermoFisher) for 5 hr and washed in Low Salt buffer (20 mM Tris-Cl pH 8.0,
693 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), High Salt buffer (identical but with 500
694 mM NaCl), LiCl buffer, and Tris-EDTA pH 8.0 plus 1 mM EDTA (TE buffer). Washes were
695 performed at 4°C for 10 min with rotation. For re-ChIP, 2% fragmented chromatin was saved as
696 input and the rest used for IP with the primary antibody (either Utx or Oct1) at 4°C overnight on a
697 rotator. Samples were then incubated with magnetic beads for 5h at 4°C. The beads were washed
698 for 10 min with Low Salt buffer, High Salt buffer, Low Salt buffer, LiCl buffer, and TE buffer
699 sequentially at 4°C. Chromatin was eluted with 300 μL IP Elution buffer at RT, then diluted 10-
700 fold in RIPA buffer. Diluted samples were then subjected to a second IP with 4 μg of the secondary
701 antibody (Oct1 or Oct4) at 4°C overnight, and incubated with magnetic beads for 5 hr at 4°C. The
702 beads were washed again as described above, then eluted with 300 μL IP Elution Buffer at RT.
703 Re-ChIP samples, together with the 2% input, were incubated at 65°C overnight to reverse
704 crosslinking. DNA was purified using phenol-chloroform-isoamyl alcohol extraction followed by
705 PCR clean up. qPCR primers can be found in Supplementary Table 6 and were confirmed to
706 generate a single product of the correct size. The results were reported as qPCR values
707 normalized to input chromatin (gDNA) and non-specific region and presented as fold enrichment.
708
709 **ChIP-seq analysis.** After chromatin was precipitated as described above, and libraries were
710 sequenced using Illumina NovaSeq. Between 22 and 26 million paired-end Illumina sequencing
711 reads were aligned to the mouse *Mm10* reference genome using Novocraft novoalign v3.8.2,
712 allowing for one random alignment of multi-mapping reads, and providing the adapter sequences
713 for automatic trimming concordant with alignment. ChIP was analyzed using the
714 MultiRepMacsChIPSeq pipeline v12.2, using options "--pe --optdist 10000 --dupfrac 0.15 --mapq
715 10 --cutoff 2 --tdep 10 --peaksize 500 --peakgap 100 --species mouse --chrskip 'chrMIPhiX' --
716 blacklist mm10.blacklist.bed".
717
718 **Immunoprecipitation.** Cells were lysed with Cell Lysis Buffer (Life Technologies) in the presence
719 of protease inhibitors (EDTA-free tablet, Roche). IP was performed using 500 µg of extract.

720 Extracts were incubated with 4 µg of anti-Utx (Cell Signaling, D3Q1I) or Oct1 (Novus Biologicals,
721 NBP2-21584) antibodies, or rabbit IgG control overnight at 4°C. Protein-antibody complexes were
722 precipitated with protein-G magnetic beads (Thremo Fisher) for 3 hr at 4°C with rotation and
723 washed 3 times with Low Salt buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1%
724 SDS, 1% Triton X-100) plus protease inhibitors. Precipitated proteins were analyzed by
725 immunoblot.
726
727 **Luciferase reporter assay.** Single or combination mutations were introduced into the Oct1 and
728 Smad consensus binding sites in the following mouse *Myog* regulatory element (*Mm10*
729 chr1:134,285,666-134,285,750). IDT g-blocks® were synthesized to contain WT sequences or
730 single or combined mutations in the Oct1 or Smad binding sites, fused upstream of the CMV basal
731 promoter (-216-13 relative to TSS). G-blocks were inserted using sequence- and ligase-
732 independent cloning (Li and Elledge, 2012) upstream of the coding sequence for a secreted nano-
733 luciferase following digestion of vector pNL2.3 (Promega) using *Eco*RV and *Hind*III. The veracity
734 of the cloned inserts was confirmed by Sanger sequencing. 200 ng of reporter plasmid were co-
735 transfected into Oct1-deficient MEFs (Shakya et al., 2009) in DMEM media lacking phenol red
736 (ThermoFisher) together with 400 ng MMP9-mCherry (Wider and Picard, 2017) in 1000 ng total
737 transfected DNA. Where indicated, 400 ng pBabePuro-Oct1 was included in the transfection mix.
738 pUC18 plasmid comprised the balance of the transfected DNA. Where indicated, transfected cells
739 were provided with 5 ng recombinant mouse TGFb1 protein (R&D Systems). mCherry
740 fluorescence was determined first by exciting at 570 nm and measuring emission at 610 nm with
741 a 100 msec time delay using an Envision Xcite Multilabel Plate Reader. Luminescence was
742 measured using Nano-Glo Luciferase (Promega) and a Modulus luminescence plate reader.

743 **Acknowledgements**

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

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

756

757 **Conflict of interest**

758 The authors declare that they have no conflicts of interest

759 **REFERENCES**

760

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

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

825 doi:10.1038/s41586-020-2493-4
826 Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI. 1995. Embryonic stem cells express
827 neuronal properties in vitro. *Dev Biol* **168**:342–357. doi:10.1006/dbio.1995.1085
828 Beddington RSP, Robertson EJ. 1989. An assessment of the developmental potential of
829 embryonic stem cells in the midgestation mouse embryo. *Development* **105**:733–737.
830 Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M,
831 Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. 2006. A bivalent
832 chromatin structure marks key developmental genes in embryonic stem cells. *Cell*
833 **125**:315–26. doi:10.1016/j.cell.2006.02.041
834 Chal J, Oginuma M, Tanoury Z Al, Gobert B, Sumara O, Hick A, Bousson F, Zidouni Y, Mursch
835 C, Moncuquet P, Tassy O, Vincent S, Miyanari A, Bera A, Garnier J, Guevara G, Hestin M,
836 Kennedy L, Hayashi S, Drayton B, Cherrier T, Gayraud-morel B, Gussoni E, Relaix F,
837 Tajbakhsh S, Pourquié O. 2015. Differentiation of pluripotent stem cells to muscle fiber to
838 model Duchenne muscular dystrophy. *Nat Biotechnol* **33**:962–9. doi:10.1038/nbt.3297
839 Cloos PAC, Christensen J, Agger K, Helin K. 2008. Erasing the methyl mark: Histone
840 demethylases at the center of cellular differentiation and disease. *Genes Dev*.
841 doi:10.1101/gad.1652908
842 Collinson A, Collier AJ, Morgan NP, Sienerth AR, Chandra T, Andrews S, Rugg-Gunn PJ. 2016.
843 Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and
844 Differentiation Defects in Human Embryonic Stem Cells. *Cell Rep*.
845 doi:10.1016/j.celrep.2016.11.032
846 Conlon FL, Lyons KM, Takaesu N, Barth KS, Kispert A, Herrmann B, Robertson EJ. 1994. A
847 primary requirement for nodal in the formation and maintenance of the primitive streak in
848 the mouse. *Trends Genet* **10**:308–308. doi:10.1016/0168-9525(94)90026-4
849 Dell'Orso S, Juan AH, Ko KD, Naz F, Perovanovic J, Gutierrez-Cruz G, Feng X, Sartorelli V.
850 2019. Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and
851 regenerative conditions. *Development* **146**. doi:10.1242/DEV.174177
852 DeVeale B, Brokhman I, Mohseni P, Babak T, Yoon C, Lin A, Onishi K, Tomilin A, Pevny L,
853 Zandstra PW, Nagy A, van der Kooy D. 2013. Oct4 is required ~E7.5 for proliferation in the
854 primitive streak. *PLoS Genet* **9**. doi:10.1371/JOURNAL.PGEN.1003957
855 Dhar SS, Lee SH, Chen K, Zhu G, Oh WK, Allton K, Gafni O, Kim YZ, Tomoiga AS, Barton MC,
856 Hanna JH, Wang Z, Li W, Lee MG. 2016a. An essential role for UTX in resolution and
857 activation of bivalent promoters. *Nucleic Acids Res* **44**:3659–3674.

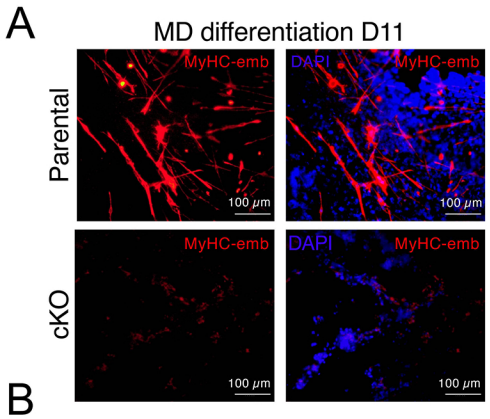
858 doi:10.1093/nar/gkv1516
859 Dhar SS, Lee SH, Chen K, Zhu G, Oh WK, Allton K, Gafni O, Kim YZ, Tomoiga AS, Barton MC,
860 Hanna JH, Wang Z, Li W, Lee MG. 2016b. An essential role for UTX in resolution and
861 activation of bivalent promoters. *Nucleic Acids Res* **44**:3659–3674.
862 doi:10.1093/nar/gkv1516
863 Diaz-Cuadros M, Wagner DE, Budjan C, Hubaud A, Tarazona OA, Donelly S, Michaut A, Al
864 Tanoury Z, Yoshioka-Kobayashi K, Niino Y, Kageyama R, Miyawaki A, Touboul J,
865 Pourquié O. 2020. In vitro characterization of the human segmentation clock. *Nature*
866 **580**:113–118. doi:10.1038/s41586-019-1885-9
867 Gnocchi VF, White RB, Ono Y, Ellis JA, Zammit PS. 2009. Further characterisation of the
868 molecular signature of quiescent and activated mouse muscle satellite cells. *PLoS One*
869 **4**:e5205. doi:10.1371/journal.pone.0005205
870 Guibentif C, Griffiths JA, Imaz-Rosshandler I, Ghazanfar S, Nichols J, Wilson V, Göttgens B,
871 Marioni JC. 2021. Diverse Routes toward Early Somites in the Mouse Embryo. *Dev Cell*
872 **56**:141-153.e6. doi:10.1016/j.devcel.2020.11.013
873 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass
874 CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-
875 Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**:576–589.
876 doi:10.1016/j.molcel.2010.05.004
877 Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. 2013.
878 Super-enhancers in the control of cell identity and disease. *Cell* **155**:934–47.
879 doi:10.1016/j.cell.2013.09.053
880 Kang J, Gemberling M, Nakamura M, Whitby FG, Handa H, Fairbrother WG, Tantin D. 2009a. A
881 general mechanism for transcription regulation by Octl and Oct4 in response to genotoxic
882 and oxidative stress. *Genes Dev* **23**:208–222. doi:10.1101/gad.1750709
883 Kang J, Gemberling M, Nakamura M, Whitby FG, Handa H, Fairbrother WG, Tantin D. 2009b. A
884 general mechanism for transcription regulation by Oct1 and Oct4 in response to genotoxic
885 and oxidative stress. *Genes Dev* **23**:208–222. doi:10.1101/gad.1750709
886 Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum
887 C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein
888 BE. 2008. Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of
889 Bivalent Domains. *PLoS Genet* **4**:e1000242. doi:10.1371/journal.pgen.1000242
890 La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti

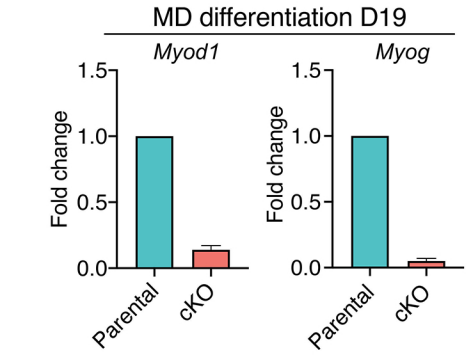
891 ME, Lönnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker
892 R, Sundström E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko P
893 V. 2018. RNA velocity of single cells. *Nature* **560**:494–498. doi:10.1038/s41586-018-0414-
894 6
895 Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. 2010. ChEA: Transcription
896 factor regulation inferred from integrating genome-wide ChIP-X experiments.
897 *Bioinformatics* **26**:2438–2444. doi:10.1093/bioinformatics/btq466
898 Li MZ, Elledge SJ. 2012. SLIC: A method for sequence- and ligation-independent cloning.
899 *Methods Mol Biol* **852**:51–59. doi:10.1007/978-1-61779-564-0_5
900 Liu Z, Lin X, Cai Z, Zhang Z, Han C, Jia S, Meng A, Wang Q. 2011. Global identification of
901 SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early
902 gastrulas. *J Biol Chem* **286**:28520–28532. doi:10.1074/jbc.M111.236307
903 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
904 RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8
905 McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G.
906 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*
907 **28**:495–501. doi:10.1038/nbt.1630
908 Meshorer E, Misteli T. 2006. Chromatin in pluripotent embryonic stem cells and differentiation.
909 *Nat Rev Mol Cell Biol* **7**:540–6. doi:10.1038/nrm1938
910 Mullen AC, Orlando DA, Newman JJ, Lovén J, Kumar RM, Bilodeau S, Reddy J, Guenther MG,
911 DeKoter RP, Young RA. 2011. Master transcription factors determine cell-type-specific
912 responses to TGF-β signaling. *Cell* **147**:565–76. doi:10.1016/j.cell.2011.08.050
913 Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Schöler H,
914 Smith A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on
915 the POU transcription factor Oct4. *Cell* **95**:379–391. doi:10.1016/S0092-8674(00)81769-9
916 Oginuma M, Moncuquet P, Xiong F, Karoly E, Chal J, Guevorkian K, Pourquié O. 2017. A
917 Gradient of Glycolytic Activity Coordinates FGF and Wnt Signaling during Elongation of the
918 Body Axis in Amniote Embryos. *Dev Cell* **40**:342-353.e10.
919 doi:10.1016/j.devcel.2017.02.001
920 Palmieri SL, Peter W, Hess H, Schöler HR. 1994. Oct-4 transcription factor is differentially
921 expressed in the mouse embryo during establishment of the firsttwo extraembryonic cell
922 lineages involved in implantation. *Dev Biol* **166**:259–267. doi:10.1006/dbio.1994.1312
923 Perovanovic J, DellOrso S, Gnochi VF, Jaiswal JK, Sartorelli V, Vigouroux C, Mamchaoui K,

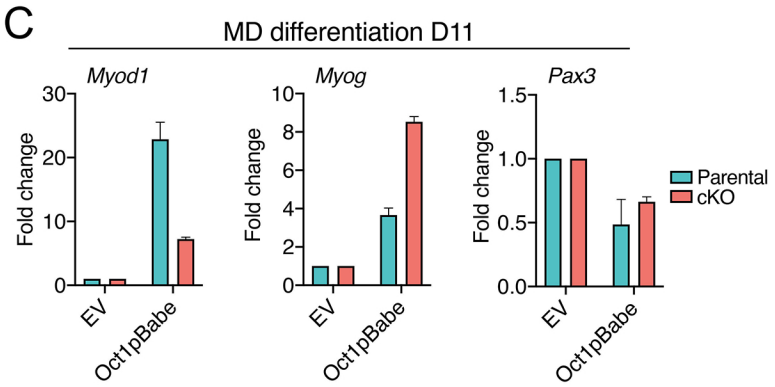
924 Mouly V, Bonne G, Hoffman EP. 2016. Laminopathies disrupt epigenomic developmental
925 programs and cell fate. *Sci Transl Med* **8**:335ra58-335ra58.
926 doi:10.1126/scitranslmed.aad4991
927 Schlesinger S, Meshorer E. 2019. Open Chromatin, Epigenetic Plasticity, and Nuclear
928 Organization in Pluripotency. *Dev Cell*. doi:10.1016/j.devcel.2019.01.003
929 Sebastiano V, Dalvai M, Gentile L, Schubart K, Sutter J, Wu GM, Tapia N, Esch D, Ju JY,
930 Hübner K, Bravo MJA, Schöler HR, Cavaleri F, Matthias P. 2010. Oct1 regulates
931 trophoblast development during early mouse embryogenesis. *Development* **137**:3551–
932 3560. doi:10.1242/dev.047027
933 Seydoux G, Braun RE. 2006. Pathway to Totipotency: Lessons from Germ Cells. *Cell*.
934 doi:10.1016/j.cell.2006.11.016
935 Shakya A, Callister C, Goren A, Yosef N, Garg N, Khoddami V, Nix D, Regev A, Tantin D. 2015.
936 Pluripotency transcription factor Oct4 mediates stepwise nucleosome demethylation and
937 depletion. *Mol Cell Biol* **35**:1014–25. doi:10.1128/MCB.01105-14
938 Shakya A, Cooksey R, Cox JE, Wang V, McClain DA, Tantin D. 2009. Oct1 loss of function
939 induces a coordinate metabolic shift that opposes tumorigenicity. *Nat Cell Biol* **11**:320–327.
940 doi:10.1038/ncb1840
941 Shakya A, Kang J, Chumley J, Williams MA, Tantin D. 2011. Oct1 is a switchable, bipotential
942 stabilizer of repressed and inducible transcriptional states. *J Biol Chem* **286**:450–9.
943 doi:10.1074/jbc.M110.174045
944 Shen Z, Kang J, Shakya A, Tabaka M, Jarboe EA, Regev A, Tantin D. 2017. Enforcement of
945 developmental lineage specificity by transcription factor Oct1. *Elife* **6**.
946 doi:10.7554/eLife.20937
947 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M,
948 Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-
949 1902.e21. doi:10.1016/j.cell.2019.05.031
950 Takahashi K, Yamanaka S. 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic
951 and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**:663–676.
952 doi:10.1016/j.cell.2006.07.024
953 Tantin D. 2013. Oct transcription factors in development and stem cells: insights and
954 mechanisms. *Development* **140**:2857–2866. doi:10.1242/dev.095927
955 Tantin D, Schild-Poulter C, Wang V, Haché RJG, Sharp PA. 2005. The Octamer Binding
956 Transcription Factor Oct-1 Is a Stress Sensor. *Cancer Res* **65**:10750–10758.

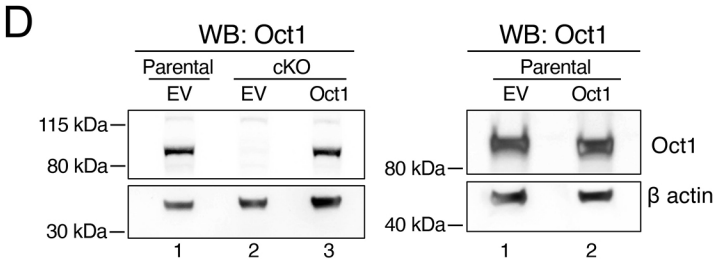
957 doi:10.1158/0008-5472.CAN-05-2399
958 Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ,
959 Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are
960 revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381–386.
961 doi:10.1038/nbt.2859
962 Wang C, Lee JE, Cho YW, Xiao Y, Jin Q, Liu C, Ge K. 2012. UTX regulates mesoderm
963 differentiation of embryonic stem cells independent of H3K27 demethylase activity. *Proc*
964 *Natl Acad Sci U S A* **109**:15324–15329. doi:10.1073/pnas.1204166109
965 Wang VEH, Schmidt T, Chen J, Sharp PA, Tantin D. 2004. Embryonic lethality, decreased
966 erythropoiesis, and defective octamer-dependent promoter activation in Oct-1-deficient
967 mice. *Mol Cell Biol* **24**:1022–32. 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

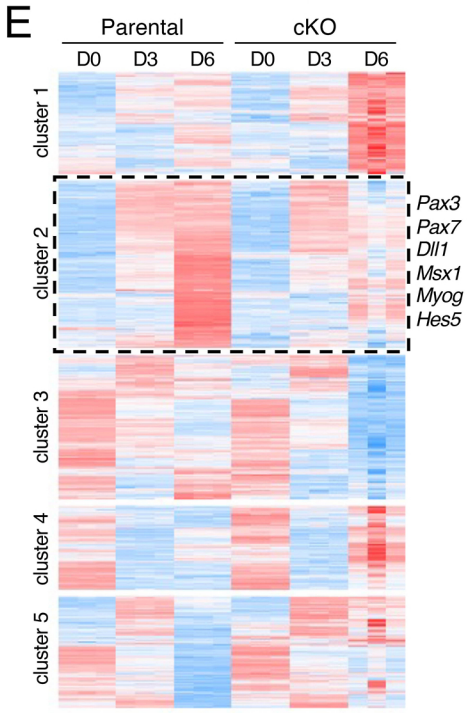
Perovanovic *et al.* Figure 1











F

Cluster 1 Enrichment

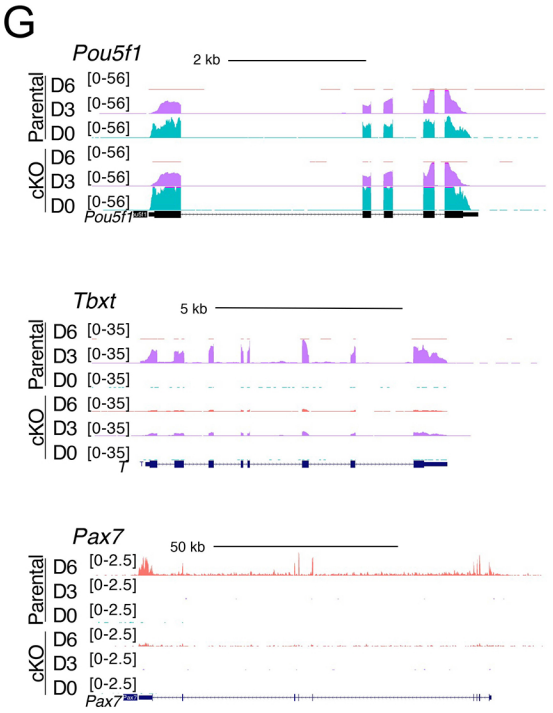
Jensen Tissue	Adj. *p*-value
Germ cell	2.8E-6
Primordium	1.5E-5

Cluster 2 Enrichment

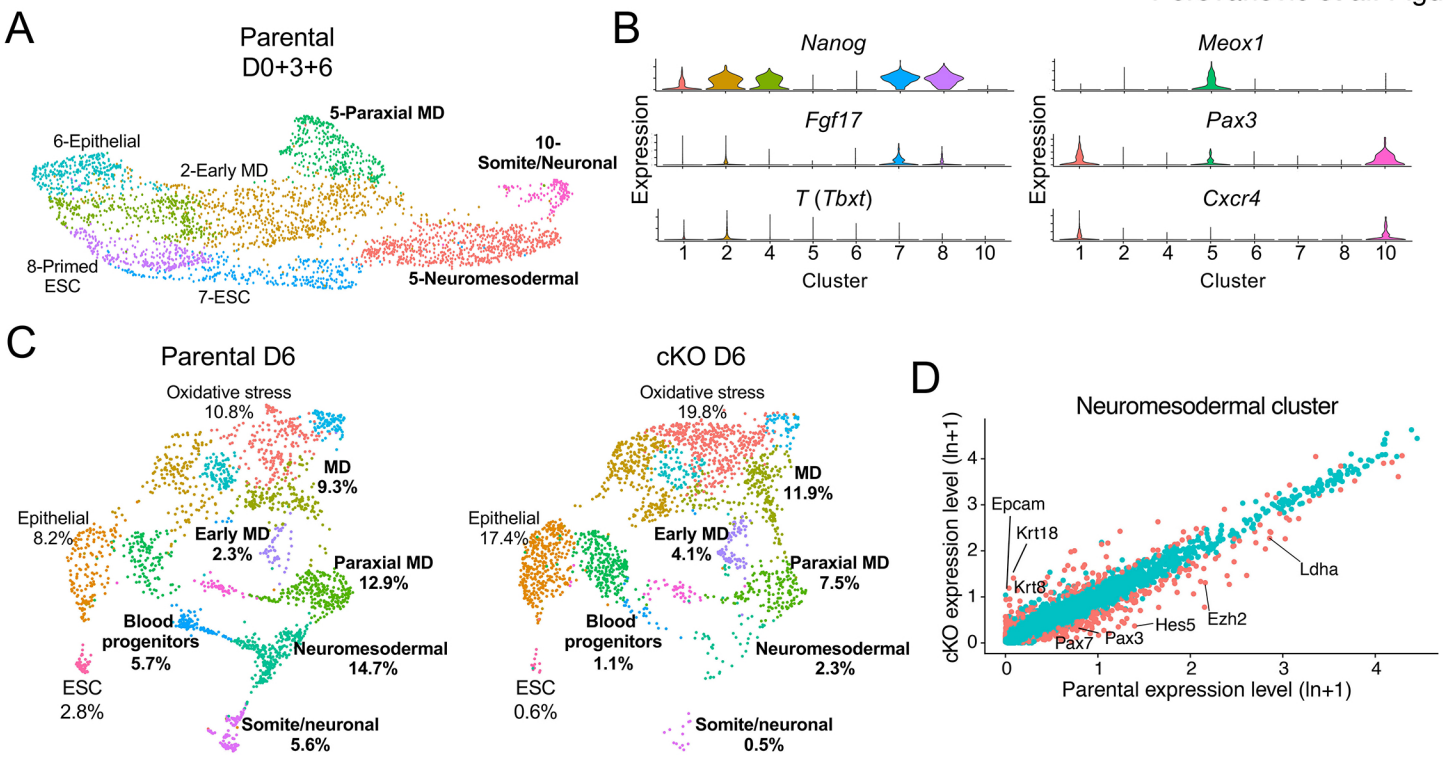
Jensen Tissue	Adj. *p*-value
Neural Tube	2.8E-6
Somite	3.8E-5

Cluster 2 Regulation

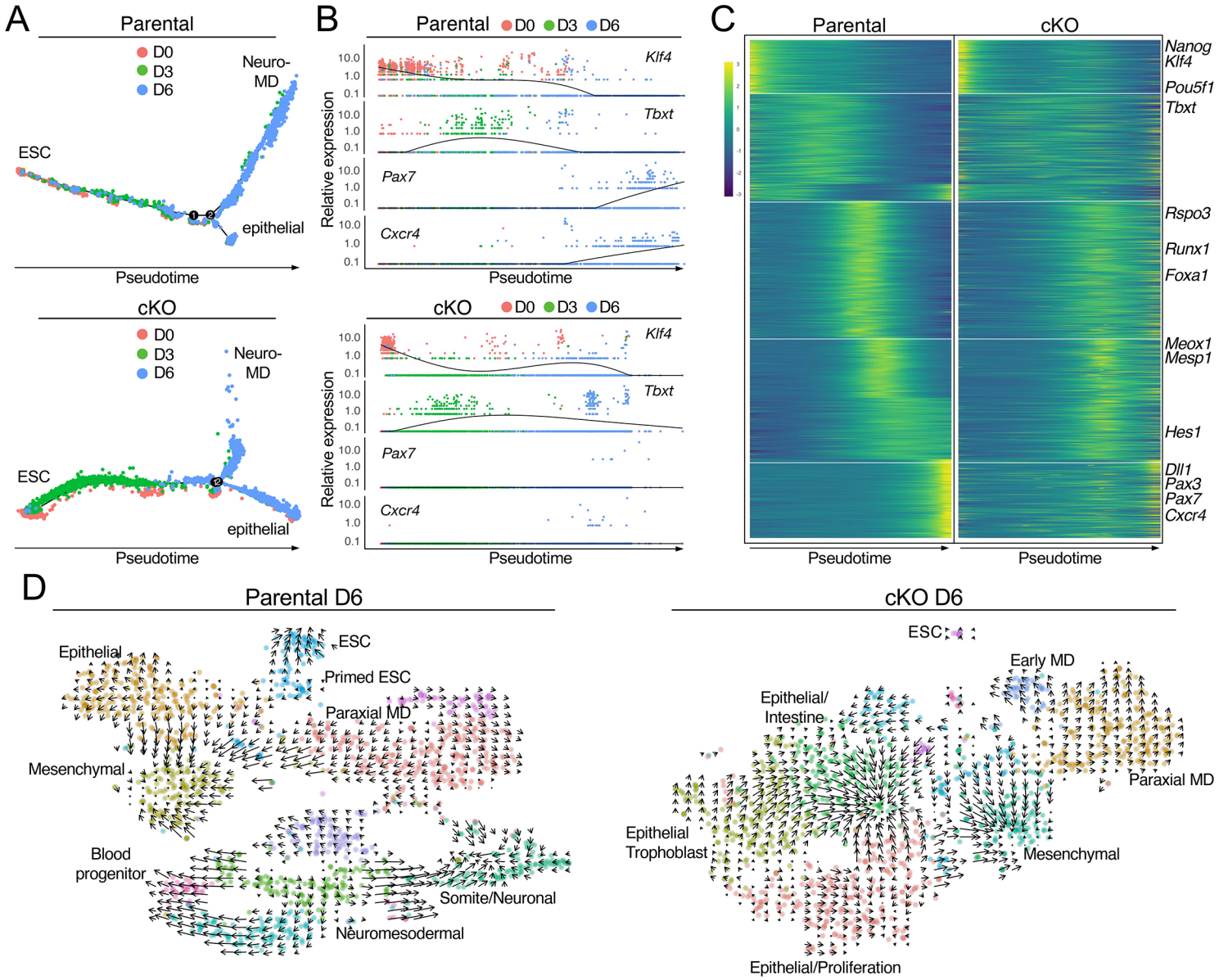
ChEA 2016	Adj. *p*-value
SUZ12	1.836E-21
MTF2	7.127E-21
JARID2	2.807E-20
EZH2	7.379E-20



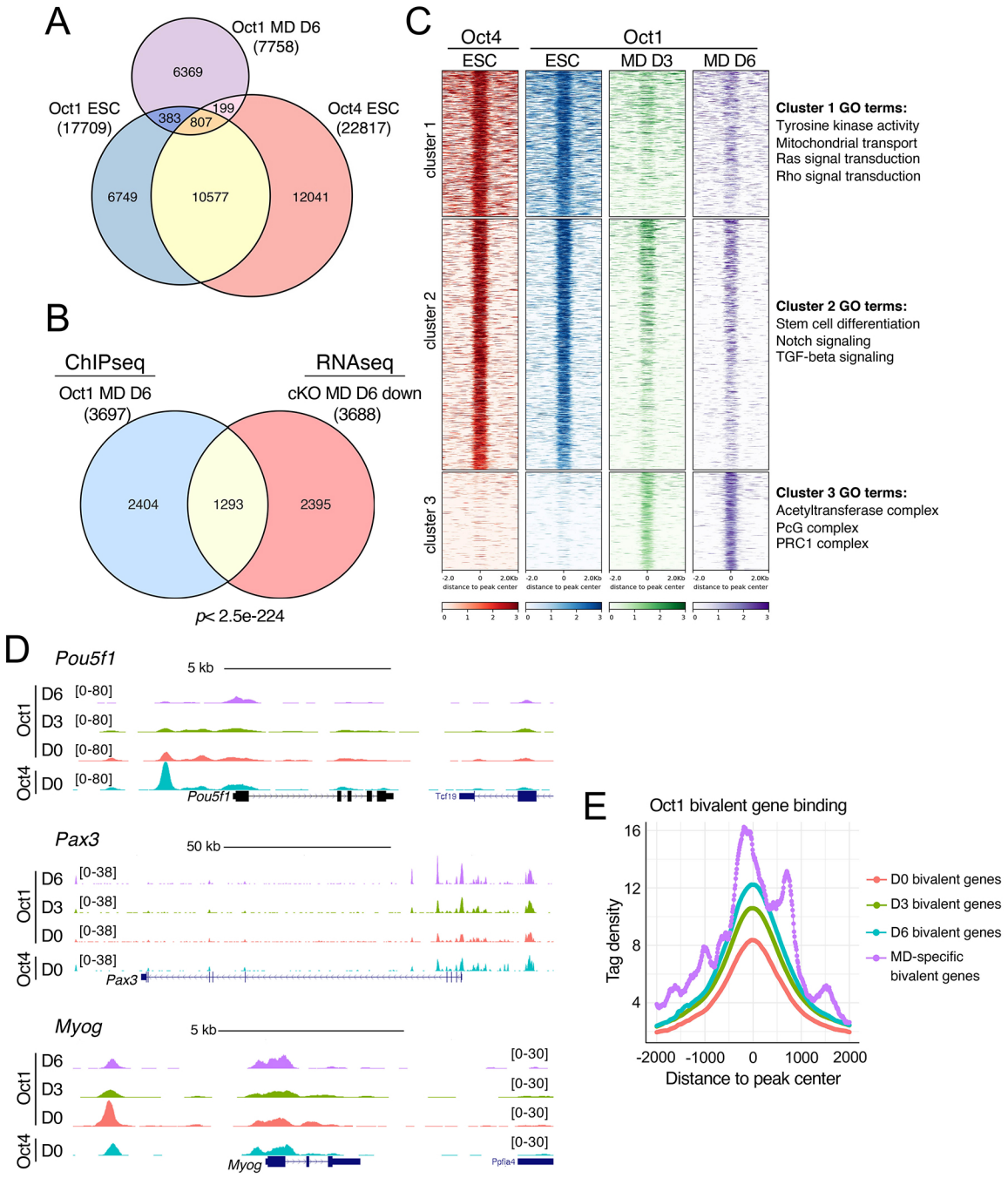
Perovanovic *et al.* Figure 2



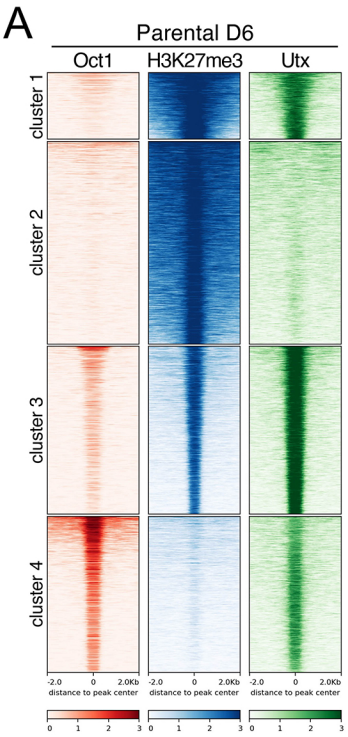
Perovanovic *et al.* Figure 3



Perovanovic *et al.* Figure 4



Perovanovic *et al.* Figure 5



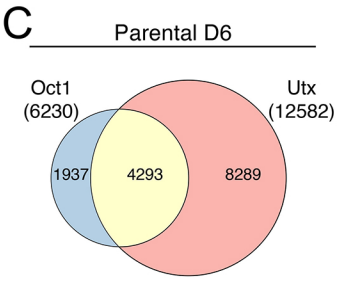
B

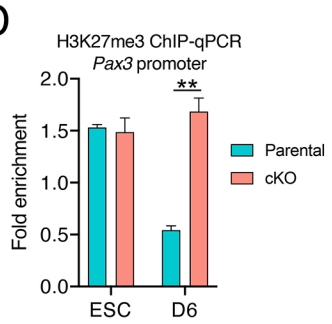
Cluster 3 Enrichment

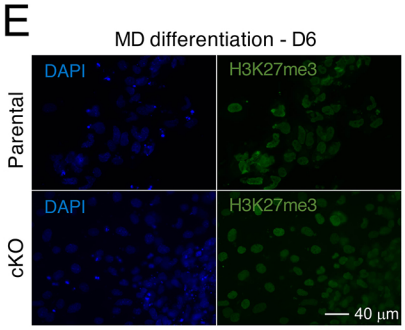
Human Phenotype Ontology	Binome FDR
Abnormality of the musculature	2.00E-46
Abnormality of muscle physiology	1.20E-44
Abnormality of skeleton morphology	7.51E-41

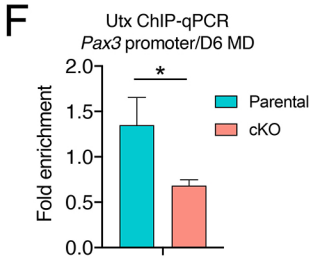
Cluster 4 Enrichment

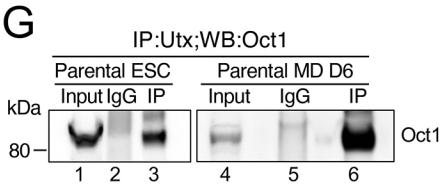
Human Phenotype Ontology	Binome FDR
Abnormality of the musculature	1.99E-39
Global developmental delay	1.72E-36
Neurodevelopmental delay	2.82E-35
Abnormality of metabolism	5.28E-35

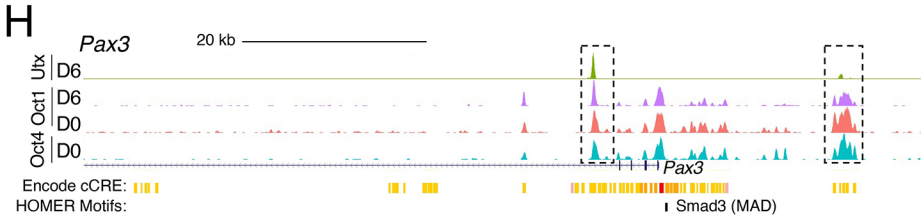




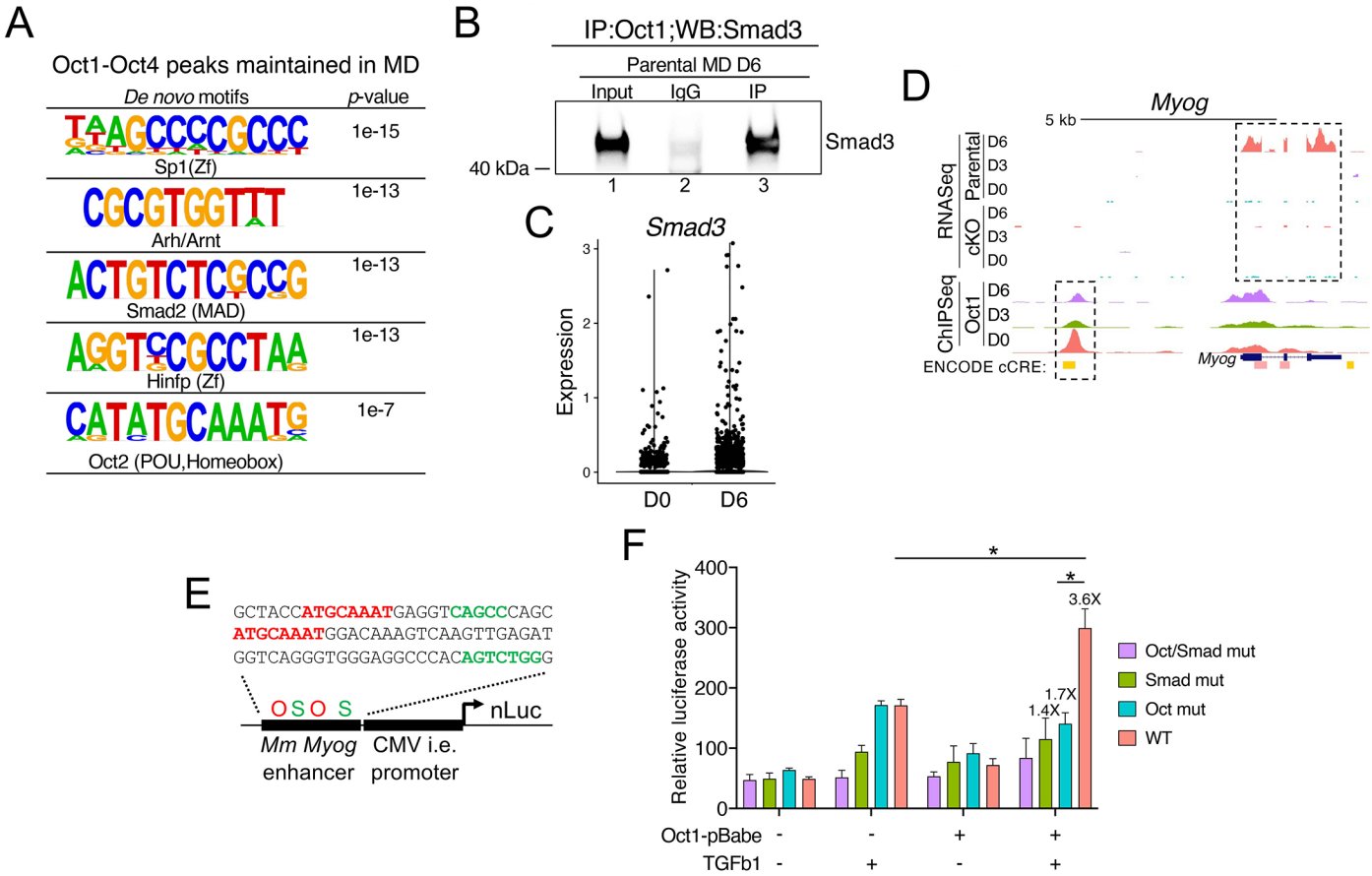
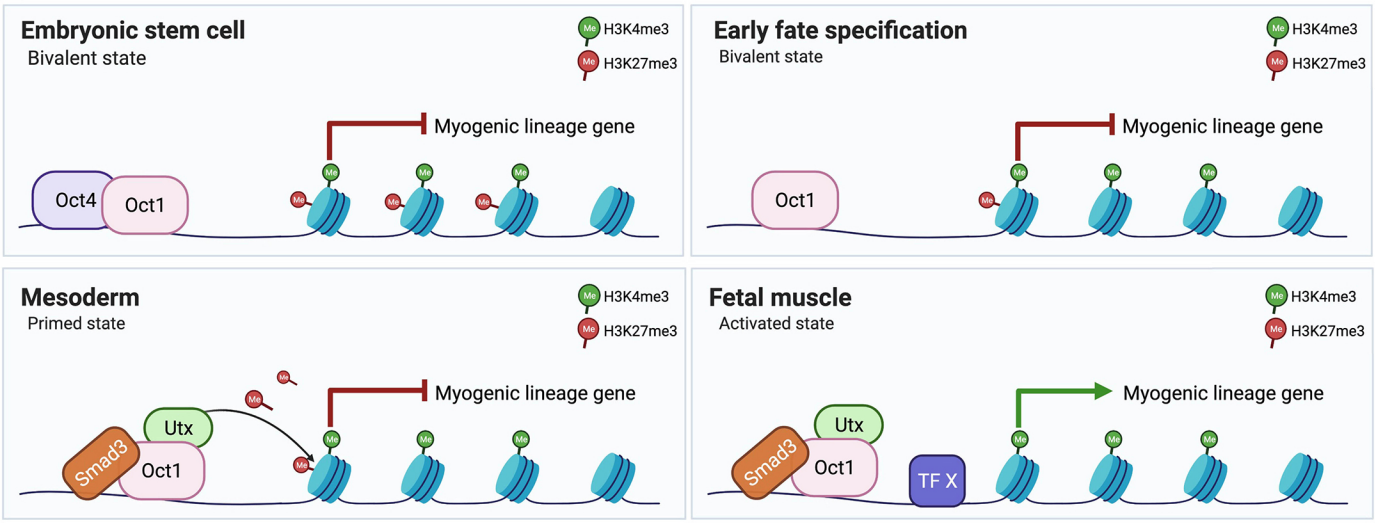








Perovanovic *et al.* Figure 6



Perovanovic *et al.* Figure 7