Oct1 recruits the histone lysine demethylase Utx to promote lineage specification

Jelena Perovanovic[1,4], Yifan Wu[1,4], Hosiana Abewe[2,4], Zuolian Shen[1,4], Erik Hughes[1,4], Jason Gertz[2,4], Mahesh B. Chandrasekharan[3,4], Dean Tantin[1,4,*]

[1]Department of Pathology
[2]Department of Oncological Sciences
[3]Department of Radiation Oncology
[4]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; Hi-ChIP; Oct1/Pou2f1; Utx/Kmd3a; Smad3

Characters in title (including spaces): 81
Words in abstract: 234
Words main text: 10,078
Character count (including spaces): 68,971
Character count (excluding spaces): 58,969
Total main figures and tables: 7

SUMMARY

The pathways used by cells to transition between pluripotent 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. We show that mesoderm-specific genes are not correctly induced early in the differentiation timecourse. Oct1-deficient cells lose temporal coherence in the induction of lineage-specific genes and show inappropriate developmental lineage branching, resulting in poorly differentiated cells states retaining epithelial characteristics. In embryonic stem cells, Oct1 co-binds with Oct4 to genes critical for mesoderm induction, and continues to bind these genes during differentiation. Oct1 binding events are enriched at the termini of chromatin loops, including loops gained with differentiation. 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, and that Oct1 and Smad3 act cooperatively at the *Myog* enhancer. Overall, these results identify Oct1 as a key mediator of the induction of mesoderm lineage-specific genes.

54 **INTRODUCTION**

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

66 The transcriptional changes underlying lineage specification require extensive chromatin
67 remodeling and spatiotemporal activation of genes encoding developmental master regulators.
68 Remodeling is potentiated by a unique chromatin landscape in pluripotent cells. Chromatin in
69 ESCs is largely accessible and lacks stable heterochromatin domains (Meshorer and Misteli,
70 2006; Schlesinger and Meshorer, 2019). A large number of genes encoding lineage-specific
71 developmental regulators are marked at promoters and gene bodies by covalently modified
72 nucleosomes that simultaneously convey activating (H3K4me3) and repressing (H3K27me3)
73 potential (Bernstein et al., 2006; Ku et al., 2008). In ESCs, these "bivalent" genes are silent or
74 expressed at low levels, but lack DNA methylation and are poised for activation. During
75 development, gene bivalency is thought to resolve via either removal of activating marks and gene
76 silencing, or removal of repressive marks and gene activation. Which poised genes resolve in
77 which direction is lineage-specific, resulting in distinct, cell fate-appropriate transcription programs
78 and durable repression of lineage-inappropriate genes. However, the programs that instruct
79 activation of the correct genes are poorly understood.

80 POU transcription factors play central roles in development (Tantin, 2013). The POU factor
81 Oct4 (encoded by *Pou5f1*) is a master regulator of the induction and maintenance of pluripotency
82 (Nichols et al., 1998; Palmieri et al., 1994; Takahashi and Yamanaka, 2006). Oct4 associates with
83 bivalent genes in ESCs (Bernstein et al., 2006), but is silenced in their differentiated progeny
84 before bivalency resolution and the induction of tissue- and cell type-specific gene expression
85 (DeVeale et al., 2013). A second POU protein, Oct1/Pou2f1, is co-expressed with Oct4, but unlike
86 Oct4 is expressed beyond pluripotency (Shen et al., 2017). Oct1 is widely expressed and required

87 for placental and embryonic development (Sebastiano et al., 2010; Wang et al., 2004).
88 Circumventing the placental defects via tetraploid complementation results in developmental
89 arrest at E8.25 with no more than five somites (Sebastiano et al., 2010). Oct1-deficient ESCs are
90 morphologically normal and proliferate and self-renewal normally, but upon differentiation show
91 decreased developmental lineage-specific gene expression and elevated expression of
92 developmentally incorrect genes, including extra-embryonic genes (Shen et al., 2017). Extended
93 incubation of these cells supports an interpretation that there is a failure to differentiate rather
94 than a kinetic delay (Shen et al., 2017). The underlying molecular mechanisms by which Oct1
95 regulates lineage differentiation are unknown.

96 Here, using Oct1 inducible-conditional ESCs, we show that pluripotent cells lacking Oct1
97 fail to form terminally differentiated myotubes. Differentiation is restored by Oct1 retroviral
98 complementation. Bulk RNA-seq profiling early in the differentiation timecourse identifies gene
99 expression abnormalities early in MD differentiation that predict the later phenotypic defects.
100 Single-cell RNA sequencing reveals that cells lacking Oct1 accumulate abnormal cell populations
101 associated with epithelial characteristics, oxidative stress and early MD, while almost completely
102 failing to reach somite-stage differentiation. Pseudotime analysis reveal increased predilection to
103 proceed down incorrect developmental trajectories, and loss of coherence in differentiation gene
104 expression programs. Oct1 and Oct4 ChIP-seq identifies co-binding of Oct1 with Oct4 in
105 pluripotent cells that is carried forward into differentiation after Oct4 is lost. H3K27Ac Hi-ChIP
106 reveals that a large number of new topological chromatin interactions gained with differentiation
107 correspond to sites of Oct1 binding. We show that Oct1 interacts with components of the Utx
108 (Kdm3a) H3K27me3 demethylase complex, and recruits Utx to developmentally appropriate
109 target genes, resulting in removal of H3K27me3 and de-repression at tested targets. Binding sites
110 for Oct1 and Oct4 in pluripotent cells in which Oct1 binding is carried forward during MD
111 differentiation are also enriched in sites for Smad transcription factors, key mediators of MD
112 specification. Oct1 and Smad3 interact and cooperatively activate gene expression in
113 differentiating cells, providing a means of restricting Oct1's activating potential to MD-specific
114 genes. The cumulative results support a model in which Oct1 "canalizes" differentiation by
115 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 subjected 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") to an MD differentiation protocol that begins to generate 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 and cKO cells were differentiated for 11 days (D11) to generate early myotubes. Immunostaining reveals robust embryonic myosin heavy chain (MyH-emb) expression in fused myotubes from parental cells, but 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 manifest defective early gene expression and differentiate into muscle poorly.

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 *Myod*, *Myog* and *Pax3* mRNA expression by RT-qPCR. Differentiating cKO ESCs transduced with Oct1 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-specific gene expression in differentiating Oct1-deficient cells

To identify how Oct1 loss perturbs MD differentiation in greater detail, we performed bulk RNA-seq using parental and Oct1 cKO cells. Three biological replicates were performed per condition.

149 ~1700, ~800 and 3,300 significantly differentially expressed genes were identified at D0, D3 and
150 D6, respectively (FDR ≤ 0.05; -1 < log2FC < 1, Table S1). Euclidean distance analysis reveals
151 tight correlations between both replicates and conditions at D0 and D3, but divergence at D6
152 relative to the other conditions, and divergence between parental and cKO D6 replicates (Figure
153 S1). Unsupervised hierarchical clustering reveals groups of genes regulated similarly between
154 parental and cKO cells, and others with differential expression (Figure 1E). Cluster 2 for example
155 shows strong induction in differentiating parental but failed induction in cKO cells at D6. Example
156 genes in this set include *Pax3*, *Pax7* and *Myog* (Figure 1E, box). GO analysis of the differentially
157 expressed genes in cluster 2 reveals association with differentiation to neural tube
158 (neuroectoderm) and somite (Figure 1F). By contrast, clusters 1 and 4, which were associated
159 with aberrant gene expression in differentiating cKO cells, were associated with lineage-
160 inappropriate terms such as primordial germ cells, vascular system and stromal cells (Figure 1F
161 and not shown). To identify potential regulators of cluster 2 genes with failed induction in
162 differentiating cKO cells, we queried the ChIP Enrichment Analysis (ChEA) database (Lachmann
163 et al., 2010). Cluster 2 genes tend to be bound by the polycomb repressor compex 2 (PRC2)
164 components Suz12 and Mtf2 (Figure 1F). Example tracks are shown in Figure 1G for *Pou5f1*
165 (pluripotency), *Tbxt* (*T*/*Brachyury*, cluster 3) and *Pax7* (cluster 2). The retained D3 expression of
166 *Pou5f1*, which encodes Oct4, in both WT and cKO cells provides a likely explanation for the good
167 correlations between D0 and D3 gene expression states, with divergence at D6 (Figure S1), as
168 the presence of Oct4 may mask loss of Oct1 at D3. These data indicate that Oct1 loss results in
169 defective induction of developmental genes that are also regulated by chromatin-modifying
170 activities that act on H3K27me3.
171
172 **Differentiating ESCs lacking Oct1 proceed down abnormal developmental trajectories**
173 To investigate perturbations in cell populations and gene expression at single-cell resolution, we
174 performed single-cell RNA-seq (scRNA-seq) using parental and cKO ESCs, and cells
175 differentiated towards MD for 3 or 6 days. Cell numbers ranged between 1000 and 2500, with
176 reads per cell ranging between 90,000 and 200,000 (see methods). Data from the different
177 conditions were subjected to graph-based clustering and uniform manifold approximation and
178 projection (UMAP) to visualize clusters of cells with similar gene expression. Integrated analysis
179 of parental cells in a pluripotent state (D0) at the two differentiation timepoints identifies a range
180 of expected cell populations, from naïve and primed pluripotency to neuromesodermal, paraxial
181 MD and early somite (Figure 2A). These populations recapitulate developmental stages

182 associated with early MD (*Fgf17* and *Tbxt*/*T*/*Brachyury*), neuromesoderm (*Hoxc9*, *Pax3*, *Sox2*),
183 paraxial MD (*Meox1*, *Twist1*, *Pax3*) and early somite and neuronal progenitors (*Cxcr4, Uncx*)
184 (Figure 2B, Figure S2 and not shown). These results provide a single cell-resolution map of early
185 MD differentiation.

186 Next, we compared the parental controls to cKO cells to identify changes in cell
187 populations and gene expression. D0 and D3 cKO cells show few differences from parental, Oct1
188 sufficient cells (Table S2). In contrast, D6 cKO cells show poor differentiation capacity with
189 reductions in key mesodermal clusters (Figure 2C). For example, populations characterized by
190 neuromesodermal and paraxial MD gene expression account for 14.7 and 12.9% of cells,
191 respectively, while populations associated with blood progenitors and somites account for 5.7%
192 and 5.6% of cells (Figure 2C, left panel). The complexity in the populations is consistent with
193 findings that somites are derived from multiple transcriptional trajectories including early paraxial
194 mesoderm and neuromesodermal progenitors (Guibentif et al., 2021). In contrast, cKO D6 cells
195 show increases in clusters that retain epithelial characteristics and dramatically decreased
196 percentages of neuromesodermal progenitors (2.3%, a six-fold decrease), paraxial MD (7.5%,
197 two-fold), blood progenitors (1.1%, >5-fold) and somites (0.5%, >10-fold, Figure 2C, right panel).
198 Gene expression comparisons between parental and cKO paraxial mesoderm clusters reveals
199 that cKO cells fail to robustly induce lineage-appropriate genes such as *Pax3* and *Pax7*, and
200 inappropriately upregulate lineage-inappropriate markers such as the epithelial-specific genes
201 *Krt8* and *Krt18* (Figure 2D). These results show that Oct1 is necessary for accurate mesoderm
202 differentiation and to suppress expression of genes for alternative lineages.

203 Pseudotime analysis of scRNA-seq data allows multiple differentiation timepoints to be
204 overlaid with defined starting (ESC) and endpoint (somite) populations. Parental control cells
205 progress through a largely linear pathway with one minor branch retaining an inappropriate
206 epithelial fate (Figure 3A, top panel). In contrast, differentiating cKO ESCs show a larger
207 proportion inappropriate branching into alternative developmental trajectories (bottom panel),
208 consistent with the diminished developmental progression of paraxial MD to the somite stage,
209 and consistent with enrichment of cells that inappropriately maintain an epithelial state (Figure 2C
210 and Figure S2). We also examined pseudotemporal gene expression using specific genes
211 associated with pluripotency and MD development. In this analysis, the position of each cell in
212 pseudotime is shown on the X-axis and the degree of expression for the indicated gene on the Y-
213 axis (Figure 3B). Parental cells show robust early expression of genes associated with
214 pluripotency such *as Klf4* that lose expression roughly mid-way through pseudotime. *Tbxt* is

215 transiently induced, while the early somite markers *Pax7* and *Cxcr4* are efficiently and
216 coordinately induced later (Figure 3B, top panel). In contrast, cKO cells exhibit prolonged
217 expression of *Klf4* and *Tbxt*, and largely fail to upregulate *Pax7* and *Cxcr4* (bottom panel). We
218 used pseudotemporal ordering to visualize expression of the 2000 most variable (dynamic) genes
219 in parental cells, and compared expression of the same genes in the same order in cKO cells
220 (Figure 3C). This analysis revealed clusters associated with pluripotency (*Klf4*, *Nanog, Pou5f1*),
221 early MD differentiation (*T*), paraxial mesoderm (*Rspo3, Meox1*), and early somites (*Dll1, Pax3*).
222 The pseudotime gene expression pattern for cKO cells revealed largely normal expression of
223 early gene expression markers, but loss of temporal coherence at subsequent steps. These
224 "fuzzy" gene expression patterns begin early in differentiation, with for example prolonged
225 expression of the MD master regulator *Tbxt* and noncoherent expression of genes associated
226 with paraxial mesoderm such as *Meox1* and *Rspo3*. The diffuse pattern is different from what
227 would be expected with a kinetic delay, which would cause a rightward shift. In addition, the
228 induction of somitic genes normally expressed later in development such as *Dll1*, *Pax3* and *Pax7*
229 was poor in cKO cells (Figure 3C).
230 RNA velocity analysis allows developmental directionality to be inferred relative to other
231 cells by comparing the ratio of unspliced, newly-synthesized pre-mRNA to mature spliced mRNA.
232 A vector is then assigned to each cell that indicates developmental direction and magnitude
233 relative to the other cells. We clustered the cells separately by genotype, resulting in different
234 patterns compared to cells clustered together (Figure 3D). Applying velocity to this clustering, we
235 found that D6 differentiated parental cells show largely ordered and canonical developmental
236 progression, e.g. paraxial MD and somite (Figure 3D, marked by long arrows). cKO cells by
237 contrast are marked by convergent/stationary epithelial profiles indicative of failed differentiation
238 potential, with cells progressing primarily towards a poorly differentiated state characterized by
239 multiple lineage markers such as intestine and trophoblast (Figure 3D, left side epithelial clusters).
240 Cumulatively, the data demonstrate that Oct1 is required for efficient early adoption of the MD
241 differentiation program, with Oct1-deficient cells unable to canalize appropriate lineage
242 specification programs.
243
244 **Oct1 occupies developmental genes in ESCs**
245 One model that explains the above data is that Oct1 occupies developmental-specific targets
246 bound by Oct4 in ESCs to mediate induction of developmentally appropriate genes and
247 repression of genes specific to alternative lineages. To begin to test this hypothesis, we performed

248 chromatin immunoprecipitation (ChIP-seq) using antibodies against Oct1 and Oct4. Prior work
249 showed that Oct1 occupancy at Oct4 targets increases during retinoic acid (RA)-mediated
250 differentiation when Oct4 expression is lost (Shen et al., 2017). RA ultimately differentiates cells
251 into neuroectodermal lineages (Bain et al., 1995). In undifferentiated ESCs, strong Oct1 binding
252 (with Oct4) was only observed in a group of ~100 genes containing Oct protein variant binding
253 sites termed MOREs (More-palindromic Octamer Recognition Element) (Shen et al., 2017). We
254 used a different Oct1 antibody with superior enrichment properties (see methods) to perform
255 ChIP-seq with undifferentiated parental Oct1-sufficient ESCs, as well as cells differentiated
256 towards MD for 3 or 6 days. Oct4 ChIP-seq in undifferentiated ESCs was performed as a parallel
257 control.

258 In pluripotent cells, ~22,000 Oct4 peaks were identified, corresponding to ~6,000 genes
259 with transcription start sites (TSS) within 20 kb (Table S3). We found that ~45% of Oct4 targets
260 directly overlap (0 bp) Oct1 peaks. Conversely ~60% of Oct1 targets directly overlap Oct4 peaks
261 (Figure 4A), indicating substantial overlap between binding of the two transcription factors. Shared
262 Oct1/Oct4 targets in ESCs include *Polr2a*, which encodes the largest subunit of RNA polymerase
263 II, *Pou5f1*, which encodes Oct4, and *Dll1*, which encodes a developmentally-inducible mediator
264 of Notch signaling expressed in the MD lineage where it regulates muscle development (Zhang
265 et al., 2021). Undifferentiated (D0) Oct4 and Oct1 binding tracks are shown in Figure S3. We also
266 confirmed Oct1 and Oct4 binding to *Polr2a*, *Pou5f1* and *Dll1* using ChIP-qPCR (Figure S3). Oct1
267 binding to *Polr2a*, which contains two adjacent MOREs that can bind four Oct proteins (Kang et
268 al., 2009a), is exceptional relative to other genes in that it is far stronger than Oct4 (100× stronger
269 for *Polr2a*, 3-10× weaker for *Pou5f1* and *Dll*). Re-ChIP (sequential ChIP) indicates that Oct1 and
270 Oct4 bind these sites simultaneously. The signal was lost in Oct1 cKO ESCs, indicating specificity
271 for Oct1 (Figure S3). Cumulatively, the data indicate that in ESCs Oct1 co-binds with Oct4 to an
272 array of targets, including developmental-specific targets.

273 We then performed Oct1 ChIP-seq using D3 and D6 MD-differentiated cells. Only 199
274 Oct4-bound peaks not occupied by Oct1 in pluripotent cells become occupied by Oct1 at D6 of
275 MD differentiation (Figure 4A). In contrast, 807 shared Oct4/Oct1 peaks in pluripotent cells
276 continue to be bound by Oct1 at D6. Analysis of these peaks using GREAT (McLean et al., 2010)
277 identifies enrichment for oxidative stress, ribosomal and mitochondrial organization, Notch
278 signaling and post-implantation development including somite formation (Figure S4A).
279 Additionally, >6000 peaks become uniquely bound by Oct1 at MD D6 (Figure 4A). >2300 peaks

280 are uniquely bound at D3 (Figure S4B). We compared Oct1 target gene binding at D6 in parental
281 cells to genes that were significantly down-regulated in cKO cells at the same timepoint (FDR <
282 0.05; negative log2(FC)), observing significant (~30%) overlap (Figure 4B).
283 To pattern Oct1 and Oct4 occupancy during MD differentiation, we applied hierarchical
284 clustering. We used a union of all peaks, as opposed to peak intersect in Figure 4A, to assess
285 varying levels of Oct1 enrichment over the course of differentiation. Three major clusters of
286 occupied targets were identified, two of which show decreased or static Oct1 binding over time
287 (Figure 4C, clusters 1 and 2). Gene ontogeny (GO) analysis indicates involvement in signaling,
288 mitochondrial function, stem cell differentiation and regulation of Notch and TGFβ signaling
289 (Figure 4C).
290 Poised developmental genes are frequently associated with simultaneous
291 H3K4me3/H3K27me3 marking of genes (Bernstein et al., 2006). We queried Oct1 binding to
292 these genes during differentiation. To identify such genes, we intersected ESC H3K27me3- and
293 H3K4me3-enriched ChIP-seq peaks from the ENCODE database. This procedure identifies 3861
294 "bivalent" genes (Table S4). We then intersected this dataset with Oct1 ChIP-seq in pluripotent
295 and MD-differentiated parental cells, observing an increase in binding over time (Figure 4D, green
296 and blue lines). A similar analysis at D6 using just MD-specific bivalent genes (generated by
297 intersecting the bivalent gene dataset with MD development GO:0007498, Table S5) shows even
298 stronger Oct1 enrichment (Figure 4D, purple line). These findings indicate that Oct1 robustly binds
299 to lineage-specific genes both in pluripotent cells and their differentiated progeny. Oct1 and Oct4
300 enrichment at three bivalent, silent and developmentally poised genes encoding MD/myogenic
301 transcription factors (*Pax3*, *Myog* and *Pax7*) is shown in Figure 4E.
302 Rather than being maintained, Oct1 binding increases with differentiation in cluster 3
303 (Figure 4C). GO analysis indicates that genes near these peaks are associated with chromatin-
304 modifying activities including histone acetyltransferase and polycomb group complexes. Oct1 and
305 Oct4 enrichment at example cluster 3 genes encoding polycomb complex members (*Ezh2*,
306 *Suz12*, *Ring1* and *Ezh1*) is shown in Figure S5A. In the scRNA-seq UMAP projection, *Ezh2*
307 expression is lower in cKO cells at D6, in particular in paraxial mesoderm (Figure S5B-C). In
308 pseudotime, Ezh1 fails to become induced later in development (Figure S5D). These data indicate
309 that during MD differentiation, Oct1 directly binds to, and mediates the induction of, genes
310 encoding epigenetic regulators associated with H3K27me3.
311
312 **Oct1 binding near nodes of chromatin loops in differentiating ESCs**

313 To understand the action of Oct1 at distal regulatory elements, we intersected Oct1 ChIP-seq
314 events in parental, Oct1-sufficient mouse ESCs that are carried forward into MD differentiation
315 with the ENCODE enhancer (H3K27Ac/K3K4me1) dataset (>3100 active ESC enhancers). Many
316 Oct1 binding events were associated with enhancers (Figure 5A). Inspecting these binding events
317 relative to their target gene transcription start sites, we also observed a gradual increase in
318 proximal promoter binding during differentiation days 3 and 6 (Figure 5B and not shown).
319 Intersecting the Oct1 binding events in ESCs with known ESC super-enhancers (Hnisz et al.,
320 2013) generated an even more significant overlap, with 47/180 super-enhancers showing Oct1
321 binding (Figure 5C). GO inspection of this set of Oct1-bound super-enhancers identifies multiple
322 terms identified with embryonic morphogenesis, as expected (Figure 5D).
323 The increase in promoter binding at D6 suggested that DNA looping may facilitate Oct1
324 interactions with transcription complexes at proximal promoters. Because no interaction data are
325 available at early steps of MD differentiation, we conducted Hi-ChIP to identify chromatin loops
326 associated with H3K27Ac genome-wide. We used pluripotent (D0) and D6-differentiated parental,
327 Oct1-sufficient cells and antibodies targeting the enhancer-enriched H3K27Ac chromatin
328 modification. Two biological replicates were performed for each condition, with between 100M
329 and 200M paired-end reads per replicate. On average, 96.3% of the reads mapped to the
330 reference *Mm10* genome (see methods). This procedure produced genome-wide H3K27Ac-
331 associated topologically-associated domains (TADs) that varied significantly between the two
332 conditions, but little between replicates (Figure S6A). 804,599 (D0+D6) chromatin loops were
333 identified, the vast majority of which were maintained after 6 days of MD differentiation (Figure
334 5E). A small number of interactions (16) were lost, however a larger number of interactions were
335 gained (490, Figure 5E).
336 Comparing the chromatin looping events to Oct1 ChIP-seq, we found that the anchor
337 regions of a very high proportion of these new interactions correspond to Oct1 binding events,
338 both for D0 and D6 (Figure 5F). Between 60-70% of the 808 differential interactions (~500)
339 corresponded to Oct1 binding events using ChIP-seq at D0 and D6. 93 (~10%) of the interactions
340 associated with Oct1 binding events were common between D0 and D6 (Table S6). Genome
341 tracks for three example genes, *Ccr4*, *Pax3* and *T* (*Tbxt*) are shown in Figure 5G. Three additional
342 genes, *Pax7*, *Cd7* and *Foxk1*, are shown in Figure S6. Cumulatively, these results identify a
343 substantial overlap between the termini of chromatin loops and Oct1 binding events early in ESC
344 differentiation to MD.
345

346 **Oct1 Recruits Utx to the *Pax3* Target Gene to Remove H3K27me3.** Lineage-appropriate
347 poised developmental genes become activated in part via the removal of repressive H3K27me3
348 marks (Bernstein et al., 2006; Dhar et al., 2016a). Our findings indicate that genes with failed
349 induction in differentiating cKO cells tend to be bound by chromatin-modifying complexes that act
350 on H3K27me3 (Figure 1F). We hypothesized that during differentiation Oct1 locally recruits
351 H3K27-specific demethylases to lineage-specific genes to mediate their induction. One such
352 demethylase is Utx/Kdm6a (Cloos et al., 2008), which has been shown to regulate MD
353 differentiation (Wang et al., 2012). To test the association between Oct1 and Utx binding, we also
354 performed ChIP-seq using H3K27me3 and Utx antibodies at differentiation day 6 using parental,
355 Oct1 sufficient cells. ~12,000 H3K27me3 peaks and ~12,000 Utx peaks were identified, each
356 corresponding to ~11,000 genes whose TSS lies within 20 kb (Table S3). Unsupervised
357 hierarchical clustering together with Oct1 peaks from the same cells identifies shared and distinct
358 peaks, including weak Oct1 binding events with strong Utx and H3K27me3 association (Figure
359 6A, cluster 3), and strong Oct1 binding events associated with weaker Utx and lack of H3K27me3
360 (cluster 4). This latter cluster includes genes induced in MD such as *Pax3, Pax7* and *Myog*. We
361 interpret cluster 3 to be developmental genes that remain poised at this timepoint, and cluster 4
362 to be lineage-appropritate genes that have been activated. GO terms associated with these
363 clusters were enriched for development and musculoskeletal abnormalities (Figure 6B).
364 Intersecting the Utx and Oct1 ChIP-seq peaks identifies a high degree of overlap, with ~70% of
365 Oct1-bound peaks also associating with Utx (Figure 6C).

366 To test if cKO cells inappropriately retain H3K27me3 at lineage-appropriate genes during
367 MD differentiation, we performed ChIP-qPCR using primers flanking a strong Oct1 peak on the
368 *Pax3* promoter. *Pax3* is induced by D6 of MD differentiation, and its expression is reduced in Oct1
369 cKO cells at this timepoint (Figure 1E and Figure 3C). As expected, H3K27me3 was robust and
370 equivalent in parental and cKO D0 undifferentiated cells. D6 parental cells showed reduced
371 H3K27me3, while in contrast, cKO cells inappropriately retained elevated H3K27me3 (Figure 6D).
372 This failure to remove H3K27me3 resulted in ~3-fold higher H3K27me3 enrichment in
373 differentiated cKO relative to parental cells. Global H3K27me3 levels are unchanged at the same
374 timepoint, as evidenced by immunofluorescence (Figure 6E). Next, we performed Utx ChIP-qPCR
375 with the same primers. Utx is enriched at the *Pax3* promoter at D6 of MD differentiation, validating
376 the ChIP-seq enrichment in parental cells. Further, Utx showed reduced enrichment in cKO cells
377 (Figure 6F). These results indicate that both Oct1 and Utx associate with *Pax3*, and that
378 differentiating Oct1-deficient cells fail to remove H3K27me3 from *Pax3*.

379 We then determined if Oct1 and Utx interact using extracts from undifferentiated and D6-
380 differentiated parental ESCs and co-immunoprecipitation. The Oct1-Utx interaction is present
381 even in ESCs, and is maintained during differentiation (Figure 6G). These results support the
382 interpretation that Oct1 interacts with Utx and recruits it to target genes during differentiation to
383 help mediate their activation. Utx recruitment by Oct1 potentially explains the failed induction of
384 lineage-appropriate genes such as *Pax3*, and the failed removal of H3K27me3 in cKO cells
385 (Figure 6D). Oct1, Oct4 and Utx binding to *Pax3* is shown in Figure 6H.
386
387 **Oct1 interacts with Smad3, and cooperates with Smad proteins at *Myog* in cells**
388 **differentiating towards MD.** The broad expression of Oct1 raised the question of how Oct1
389 mediates gene activation specifically at MD lineage-specific genes. Chromatin accessibility,
390 lineage-specific signaling and co-bound transcription regulators may provide this specificity. We
391 performed motif analysis (Heinz et al., 2010) using DNA sequences ±100 bp from the center of
392 the 807 peaks co-bound by Oct1 and Oct4 in pluripotent cells that remain Oct1-bound during MD
393 differentiation (Figure 4A). This procedure identifies not only Oct1 sites, but also binding sites for
394 Smads, the terminal transcription regulators of TGFβ/Nodal/BMP signaling (Figure 7A). TGFβ
395 signals and downstream Smad transcription factor activity are critical for MD induction (Conlon et
396 al., 1994; Zhou et al., 1993). In addition, one study identified Oct1 transcription factor motifs near
397 binding events for zebrafish Smad2/3, and further showed that ectopic expression of the zebrafish
398 Oct1 ortholog enhances mesoderm induction, that zebrafish Smad2 and Oct1 physically interact,
399 and that the two factors cooperate to enhance transcription (Liu et al., 2011). Mammalian Smad3
400 also interacts with Oct4 and co-occupies target genes in pluripotent cells (Mullen et al., 2011).
401 Consistent with these findings, Oct1 antibodies efficiently co-precipitate Smad3 in D6 MD-
402 differentiated cells (Figure 7B). *Smad3* is expressed in undifferentiated parental populations, but
403 further induced by MD differentiation (Figure 7C).
404 To gain more insight into mechanism, we tested if Oct1 and Smad proteins synergize to
405 regulate gene expression. The *Myog* gene fails to be induced in D6 differentiating Oct1 cKO ESCs
406 and contains a strong Oct1 binding event overlapping with an annotated distal enhancer ~3 kb
407 upstream of the *Myog* TSS (Figure 7D). The region is also located within a known super-enhancer
408 associated with myotubes (Hnisz et al., 2013). We cloned an 85-bp region of this enhancer
409 centered on the Oct1 ChIPseq peak together with the CMV core promoter upstream of a
410 nanoLuciferase reporter cassette (Figure 7E). This region contains two octamer sites and two

411 Smad sites, and overlaps with an ENCODE distal enhancer signature (Abascal et al., 2020). The
412 reporter vector was transfected into MEFs lacking Oct1 (Wang et al., 2004). The cells were also
413 co-transfected with a plasmid constitutively expressing secreted mCherry (Wider and Picard,
414 2017) as an internal standard (Figure 7D). To selectively express Oct1 and/or activate Smad
415 proteins, cells were co-transfected with a construct encoding mouse Oct1 and/or treated with
416 recombinant TGFβ1. The two treatments together generate cooperatively activate the reporter
417 (Figure 7F). For example, constructs containing only active Smad or Oct1 binding sites (Figure
418 7E , green and blue) displayed little more activity than constructs with no sites (purple), but the
419 two sites together generate significantly increased activity (red). The affect was attenuated in the
420 absence of supplemented Oct1 or Smad activation. These results indicate that Oct and Smad
421 sites at the *Myog* enhancer cooperate to drive gene expression.

422 **Discussion**

423 During development, chromatin transitions from a pluripotent state permissive for all embryonic
424 lineages to a lineage-restricted state. Pluripotent cells maintain genes encoding developmental-
425 specific mediators in a poised configuration that allows for later induction in lineage-appropriate
426 cell types, or stable repression in all other lineages (Bernstein et al., 2006). Oct1-deficient animals
427 manifest defective induction of MD-derived somites, cardiac tissue and blood cells (Sebastiano
428 et al., 2010; Wang et al., 2004). The co-expression of Oct1 with Oct4 in pluripotent cells and
429 maintenance of Oct1 expression beyond Oct4 silencing (Shen et al., 2017), together with the
430 severe developmental phenotype of Oct1 null animals, prompted us to study the role of Oct1
431 during differentiation into the MD lineage. Here we show that Oct1-deficient ESCs manifest
432 defective MD differentiation, including failure to generate skeletal muscle myotube structures.
433 These cells fail to express *Myod* and *Myog* mRNA, and myosin heavy chain protein. RNA-seq
434 reveals that Oct1 is essentially for induction of the MD lineage-specific gene program. Using ChIP-
435 seq, we show that Oct1 continues to bind developmental lineage-specific gene enhancers during
436 mesodermal differentiation that it co-binds with its paralog Oct4 in pluripotent cells. Oct1 binding
437 is also gained at the promoters of these genes during differentiation. H3K27me3-specific Hi-ChIP
438 reveals a net increase in topological DNA contacts, and co-localization with Oct1 binding events
439 at many of the new contact sites, indicating that Oct1 tends to utilize new DNA loop interactions
440 as a means of regulating gene expression.

441 Oct1 helps mediate the activation of many bound lineage-specific genes that are silent
442 and marked by H3K27me3 in pluripotent cells. We show that Oct1 and Utx binding events
443 significantly overlap, that Oct1 interacts with Utx and that Oct1 localizes Utx to a MD-specific
444 target to remove repressive H3K27me3 marks. Specificity is achieved at least in part through
445 cooperative interactions with Smad proteins, known drivers of MD induction. Because Oct1 is
446 widely expressed and results in defective differentiation into multiple lineages (Shen et al., 2017),
447 we anticipate that this mechanism may act widely, with the only unique aspect being the
448 association with lineage-driving transcription regulators such as Smad proteins in the case of MD.

449 A central role for Oct1 in “canalizing” differentiating pluripotent cells in early steps of
450 mesodermal specification was shown using single-cell RNA-seq with Oct1 cKO ESCs. In the
451 absence of Oct1, cells undergoing differentiation retain an epithelial state, and achieve
452 neuromesodermal and somite-stage gene expression patterns poorly and in reduced numbers.
453 Increased populations of cells appear that are associated with oxidative stress, consistent with
454 findings that cells lacking Oct1 have elevated reactive oxygen species levels (Tantin et al., 2005).

455 cKO cells mis-express developmentally inappropriate genes (e.g., *Epcam* in a neuromesodermal
456 cluster) and undergo inappropriate developmental branching towards poorly differentiated states
457 marked by epithelial gene expression and oxidative stress. Unlike loss of lineage-specific master
458 regulators, where cell fate often shifts to a specific alternate lineage, these cells lack clear lineage
459 signatures. The induction of genes important early in MD differentiation such as *Tbxt* becomes
460 weaker and loses temporal coherence. Later, there is failed induction of genes such as *Pax7* and
461 *Cxcr4*. This result indicates that Oct1 helps facilitate the lineage program specified by other
462 master regulator transcription factors.

463 During MD differentiation, Oct1 occupies a subset of genes bound by Oct4 in pluripotent
464 cells. These include lineage-specific developmental mediators and genes encoding chromatin-
465 modifying enzymes. Oct1 occupancy on these latter genes increases with differentiation,
466 suggesting a critical role in canalization of cell fate. Additionally, we find that Oct1 recruits Utx to
467 lineage-specific, Oct1-bound targets such as *Pax3*. The lack of Utx recruitment to lineage-specific
468 genes in cKO cells is consistent with the abnormal retention of H3K27me3 at the promoter and
469 the failed transcriptional upregulation. No global changes in H3K27me3 were observed,
470 consistent with recruitment by Oct1 to mediate focal removal (Wang et al., 2012). An "anti-
471 repression" mechanism for Oct1 involving the removal of repressive chromatin marks has been
472 described for Oct1 before, for H3K9me2 (Shakya et al., 2011). To our knowledge this is the first
473 description of a role for Oct1 and K3K27me3. Loss of Utx is compatible with maintainance of an
474 undifferentiated state, but interferes with MD differentiation (Dhar et al., 2016b).

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

487 components may solidify lineage specification by aiding the repression of the large number of
488 genes specific to alternative lineages (Collinson et al., 2016).

489 Under MD differentiation conditions, exogenous Oct1 increases expression of the terminal
490 differentiation markers *Myod* and *Myog*, while decreasing the early lineage-specification marker
491 *Pax3*, which is transiently expressed during MD differentiation and is necessary for later
492 expression of myogenic genes. Because *Pax3* is no longer expressed in terminally differentiating
493 cells, these results suggest that ectopic Oct1 may enable transit through a *Pax3*-expressing
494 intermediate to potentiate productive terminal differentiation. More investigation into this pathway
495 is necessary to test if Oct1 can be used to more efficiently differentiate pluripotent cells. Such
496 improvements could have implications for regenerative medicine.

497 **Figure Legends**

498

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

515

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

529

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

544

545 **Figure 4. Oct1 co-occupies target sites with Oct4 in ESCs, and continues to bind lineage-**
546 **specific MD genes during differentiation.** (**A**) Venn diagram showing common and unique Oct4
547 and Oct1 binding sites based on ChIP-seq in parental ESCs and at D3 and D6 of mesodermal
548 differentiation. (**B**) Comparison of Oct1 binding events and genes down-regulated by Oct1 loss at
549 MD differentiation day 6. Target genes (< 20 kb) were used rather than peaks, explaining the
550 differing numbers compared to panel (A). (**C**) A matrix of Oct4 and Oct1 ChIP-seq enrichment 2
551 kb around peak centers was computed for the merged peak list and shown as heatmap. Positions
552 in color show high enrichment and white shows no enrichment. Associated GO terms are shown
553 at right. (**D**) Example Oct4 and Oct1 binding to three cluster 2 bivalent genes: *Pax3*, *Myog* and
554 *Pax7*. (**E**) Oct1 enrichment based on tag density at peak center at annotated bivalent genes in
555 ESCs (red), and at MD differentiation D3 and D6 (green and blue). An additional analysis was
556 performed for MD-specific bivalent genes at MD differentiation D6 (purple).

557

558 **Figure 5. Oct1 occupies developmental enhancer sites at the nodes of topological**
559 **interactions with target genes.** (**A**) Oct1 binding events in ESCs that were carried over to either
560 D3 or D6 of MD differentiation (1241 peaks) were intersected with known active enhancer regions
561 from ENCODE. Nearly half the peaks (562/1241) were enhancer associated (0 bp overlap). (**B**)

562 Oct1 binding events were mapped to the TSS of the nearest gene and binned by distance. ESCs
563 and MD D6 are shown. (**C**) Annotated super-enhancers (Hnisz et al., 2013) were intersected with
564 Oct1 binding events in ESCs. Greater than 25% (47/180) were associated with Oct1 binding. (**D**)
565 GO terms associated with Oct1-bound super-enhancers. (**E**) A scatter plot is shown comparing
566 Log(10) fold-change in H3K27Ac Hi-ChIP interactions and mean normalized counts (2 replicates).
567 D6 differentiated to D0 (pluripotent) parental ESCs. Adj. *p*-value cutoff ≤0.1 interactions are shown
568 in red. Plot was generated using DEseq2. (**F**) Venn diagrams showing overlap between
569 interactions changed with differentiation and Oct1 binding events. The 506 differential interactions
570 in panel (E) were broken down into 808 unique anchor sites showing altered DNA looping, and
571 compared with Oct1 binding at D0. The number of overlapping events represents a 3.85-fold
572 increase over random expectation. The 502 regions at D6 represents an 8-fold increase. (**G**)
573 Comparison of Oct1 ChIP events with topological interactions. Top: group-autoscaled Oct1 ChIP-
574 seq track heights in pluripotent parental cells (D0) and after six days of MD differentiation (D6) for
575 three representative genes: *Ccr4*, *Pax3* and *T* (*Tbxt*). Bottom: *Mm10* H3K27Ac Hi-ChIP genome
576 tracks. Black lines depict 1 to 4 interactions, and red lines 5 or more. Highlighted regions show
577 areas of overlap between TAD anchor sites and Oct1 ChIP-seq binding sites. Highlighted regions
578 were centered ±1 kb over the anchor site.
579
580 **Figure 6. Oct1 recruits Utx to demethylate H3K27me3 at lineage-appropriate genes.** (**A**)
581 Oct1, H3K27me3 and Utx ChIP-seq enrichment 2 kb around peak centers was computed for the
582 merged peak list and shown as heatmap. (**B**) Human Phenotype Ontology terms for Cluster 3 and
583 4 genes from (A) are shown. (**C**) The Oct1 and Utx peak lists were intersected (overlap ≥ 1 bp)
584 and plotted as a Venn diagram. 69% of Oct1 peaks overlapped with Utx binding events. (**D**)
585 H3K27me3 ChIP-qPCR enrichment at the *Pax3* promoter in parental and cKO cells at MD
586 differentiation D0 (ESC) and D6. Normalized fold-enrichment is shown relative to both an isotype
587 control antibody and to a nonspecific genomic region encoding the 28S ribosomal subunit. Data
588 represent an average of 3 biological replicates. Error bars depict ±SEM. (**E**) H3K27me3
589 immunofluorescence images from D6 MD-differentiated parental and cKO cultures. Images were
590 collected a 40× magnification. (**F**) Utx enrichment at the *Pax3* promoter in parental and cKO cells
591 on differentiation D6. Fold-enrichment relative to an isotype control antibody and a nonspecific
592 28S region is shown. Data represent an average of 3 independent biological replicates. Error bars
593 depict ±SEM. (**G**) Utx immunoprecipitation followed by Oct1 immunoblot using parental ESCs, or

594 ESCs MD-differentiated for 6D. (**H**) Signal tracks (*Mm10* v.D191020) showing Oct4, Oct1 and Utx
595 enrichment at the *Pax3* locus 5' region. Y-axes were group-autoscaled for each gene. Positions
596 of identified HOMER motifs are shown below.
597
598 **Figure 7. Oct1 and Smad3 cooperate to drive expression of mesoderm-specific genes.** (**A**)
599 HOMER motif analysis of Oct1 peaks that are shared with Oct4 in ESCs, and maintained after D6
600 of MD differentiation. (**B**) Smad3 immunoblot using cell lysates immunoprecipitated with Oct1
601 antibodies, or rabbit IgG controls. D6 MD-differentiated cells were used. 20% input is shown (lane
602 1). (**C**) *Smad3* gene expression is shown in violin plots for parental ESCs at D6 of MD
603 differentiation. (**D**) Signal tracks (*Mm10* v. D191020) showing Oct1ChIP-seq enrichment at the
604 *Myog* locus. Shown above are RNA-seq tracks in differentiated and undifferentiated parental and
605 cKO cells. ENCODE-annotated regulatory elements are shown below. Pink: proximal promoter.
606 Yellow: distal enhancer. (**E**) Scheme for Oct1/Smad reporter assay. A segment of a mouse *Myog*
607 enhancer element containing multiple Oct and Smad motifs was cloned with the core CMV
608 promoter upstream of secreted nLuc and co-transfected into Oct1-deficient MEFs together with a
609 construct encoding constitutive secreted mCherry as a normalization control. Added TGFb1 and
610 co-transfected mouse Oct1 supply Oct1 and Smad3 activity. (**F**) Transfected WT or Oct/Smad
611 mutant *Myog* enhancer constructs were supplied with Oct1, recombinant purified TGFb1
612 treatment, or both. Secreted luciferase activity was assessed relative to secreted mCherry
613 expressed from a co-transfected plasmid. An average of experimental triplicates is shown. Error
614 bars denote ±SEM. For the situation in which both Oct1 and TGFb1 are both supplied, fold
615 changes relative to a double-mutant construct are also shown. (**G**) Sequential model of bivalency
616 resolution for lineage-appropriate (MD-specific) genes. Pluripotent cells co-express Oct1 and
617 Oct4, which co-bind to poised targets. Upon loss of pluripotency and Oct4, Oct1 continues to
618 occupy these genes. TGFβ signals allow co-binding of Oct1 and Smad proteins to MD-specific
619 targets, recruitment of Utx and demethylation of H3K27me3. Later, other transcription factors
620 serve as primary "on" switches for muscle-specific gene expression.

621 **Methods**

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

628

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

633

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

646

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

651

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

654 empty vector control. The vector was co-transfected together with the pCL-Eco packaging plasmid
655 into HEK293T cells to generate retroviral particles. Retroviral supernatants were collected 48 hr
656 later, filtered through 0.45 μm filters and applied to ESCs cultures maintained on puromycin-
657 resistant feeder fibroblasts (ThermoFisher). The mixed population of cells was subjected to
658 selection with puromycin for 48 hr.
659
660 **RT-qPCR.** RNA was isolated using TRIzol (ThermoFisher). cDNA was synthesized using a
661 SuperScript Vilo cDNA Synthesis Kit (ThermoFisher). RT-qPCR oligonucleotide primers are listed
662 in Table S7 and were confirmed to generate a single product of the correct size. To ensure specific
663 PCR amplification, every RT-qPCR run was followed by a dissociation phase analysis
664 (denaturation curve) to confirm the presence of single amplified peaks.
665
666 **Bulk RNA-seq.** RNA was prepared from three independent cultures of undifferentiated or 3 d and
667 6 d MD-differentiated parental or cKO ESCs. Poly(A) RNA was purified from total RNA samples
668 (100–500 ng) with oligo(dT) magnetic beads, and stranded mRNA sequencing libraries were
669 prepared as described using the Illumina TruSeq mRNA library preparation kit and RNA UD
670 Indexes. Molarity of adapter-modified molecules was defined by qPCR using the Kapa
671 Biosystems Library Quant Kit. Individual libraries were normalized to 1.3 nM. Sequencing libraries
672 were chemically denatured and applied to an Illumina NovaSeq flow cell using the NovaSeq XP
673 chemistry workflow. Following transfer of the flowcell to an Illumina NovaSeq instrument, 2×51
674 cycle paired-end sequence was performed using a NovaSeq S1 reagent kit. Between 13 and 18
675 million paired-end reads were generated for each condition. More than 99% of aligned reads
676 mapping to the correct strand.
677
678 **Bulk RNA-seq analysis.** The Illumina adapter sequence was trimmed using cutadapt version
679 1.16. Fastq data quality were checked using Fastqc verision 0.11.5. Reads were aligned to the
680 mouse *Mm10* genome using STAR version 2.7.3a in two-pass mode. Aligned reads were checked
681 for quality using the Picard tools' CollectRnaSeqMetrics command to count the number of read-
682 matching exons, UTRs, introns and intergenic regions, and to calculate normalized gene
683 coverage from the top 1000 expressed transcripts. Between 13 and 18 million paired-end reads
684 were generated for each condition, with >99% of aligned reads mapping to the correct strand.
685 Differentially expressed genes were identified using a 5% FDR with DESeq2 version 1.24.0 (Love

686 et al., 2014). Genes with a count of at least 50 in one or more samples were tested. Genes
687 showing at least 2.5-fold change of expression at an adjusted $p<0.01$ were selected as
688 differentially expressed. Figures were generated in R version 4.0.0 using functions from ggplots
689 libraries and pheatmap.
690
691 **Single-cell RNA-seq.** Single cell transcriptomes were analyzed as described previously
692 (Dell'Orso et al., 2019). The 10X Genomics Chromium Single Cell Gene Expression Solution with
693 3' chemistry, version 3 (PN-1000075) was used to tag individual cells with 16 bp barcodes and
694 specific transcripts with 10 bp Unique Molecular Identifiers (UMIs) according to manufacturer
695 instructions. Briefly, single-cell suspensions were isolated using trypsinization and resuspension
696 in PBS with 0.04% BSA (ThermoFisher). Suspensions were filtered through 40 μm cell strainers.
697 Viability and cell count were assessed using a Countess I (ThermoFisher). Equilibrium to targeted
698 cell recovery of 6,000 cells along with Gel Beads and reverse transcription reagents were loaded
699 to Chromium Single Cell A to form Gel-bead-in Emulsions (GEMs). Within individual GEMs, cDNA
700 generated from captured and barcoded mRNA was synthesized by reverse transcription at 53°C
701 for 45 min. Samples were then heated to 85°C for 5 min. Subsequent A tailing, end repair, adaptor
702 ligation and sample indexing were performed in bulk according to manufacturer instructions. The
703 resulting barcoding libraries were qualified on Agilent Technology 2200 TapeStation system and
704 subjected to qPCR using a KAPA Biosystems Library Quantification Kit for Illumina Platforms
705 (KK4842). The multiple libraries were then normalized and sequenced on an Illumina NovaSeq
706 6000 using the 2×150 PE mode.
707
708 **Single-cell RNA-seq data processing and clustering.** Sequences from the Chromium platform
709 were de-multiplexed and aligned using CellRanger ver. 3.1.0 (10X Genomics) with default
710 parameters mm10-3.0.0. Clustering, filtering, variable gene selection and dimensionality
711 reduction were performed using Seurat ver.3.1.5 (Stuart et al., 2019) according to the following
712 workflow: 1, Cells with <300 detected genes and >10000 genes were excluded further analysis.
713 2, Cells with <12% UMIs mapping to mitochondrial genes were retained for downstream analysis.
714 3, The UMI counts per ten thousand were log-normalized for each cell using the natural logarithm.
715 4, Variable genes (2000 features) were selected using the FindVariableFeatures function. 5,
716 Common anchors between the three parental timepoints (Figure 2A) or parental and cKO D6
717 (Figure 2C) were identified using FindIntegrationAnchors function that were further used to

718 integrate these sets. 6, Gene expression levels in the integrated set were scaled along each gene
719 and linear dimensional reduction was performed. The number of principal components was
720 decided through the assessment of statistical plots (JackStrawPlot and ElbowPlot). 7, Cells were
721 clustered using a by a shared nearest neighbor (SNN) modularity optimization-based clustering
722 algorithm and visualized using two-dimensional uniform manifold approximation and projection
723 (UMAP). 8, Cluster identities were defined based on the distribution of the specific markers.
724 Differentiational gene expression analysis between the parental and cKO clusters was performed
725 using FindMarkers. Genes with adjusted $p<0.01$ were marked red on scatter plots. Using this
726 analysis, the cell numbers were between 1000 and 2500. Reads/cell ranged between 90,000 and
727 200,000. Genes/cell ranged between 2,300 and 7,000. The percentage of reads mapping to
728 reference mouse genome was 90% or higher, and the reads confidently mapping to exonic
729 regions was 70% or higher.
730
731 **Pseudotime and Velocity analysis.** Trajectory analysis of scRNA-seq was performed using
732 Monocle v.2.16.0 (Trapnell et al., 2014). Parental and cKO sets were filtered using the same
733 parameters as above and merged to generate WT and cKO sets. Cells were ordered based on
734 gene lists for the ESC (beginning) and somite (end) clusters in parental UMAP (Figure 2A). Next,
735 we performed dimensional reduction using the DDRTree method to visualize the dataset, ordered
736 the cells by global gene expression levels, and visualized the trajectory of the dataset. Veolocity
737 analysis was performed using the velocyto package (La Manno et al., 2018). Loom files were
738 produced using following paramters: velocyto run10x -m mm10. rmsk.gtf genes.gtf. Gtf files were
739 produced from the Cell Ranger pipeline. Velocity embeddings were produced using the velocyto.r
740 and SeuratWrappers packages. Matrices were filtered using following parameters:
741 nFeature_spliced > 300, nFeature_spliced < 10000, nFeature_unspliced > 200,
742 nFeature_unspliced < 6000, percent.mt < 12. Velocity was calculated using RunVelocity using
743 following paremeters: deltaT = 1, kCells = 25, fit.quantile = 0.02. Velocity embedding were
744 projected on T-SNE maps using the show.velocity.on.embedding.cor function.
745
746 **ChIP.** ChIP-qPCR and ChIP-seq were performed as previously described (Perovanovic et al.,
747 2016). Briefly, WT and cKO cells were crosslinked with 1% formaldehyde for 10 min and
748 quenched for 5 min using 2.5M glycine. Culture plates were washed using ice cold PBS and cells
749 were harvested by cell scaping. Cells were lysed in Farnham buffer (5 mM Pipes pH 8.0, 85 mM
750 KCl, 0.5% NP-40) and subsequently in RIPA buffer (phosphate-buffered saline, 1% NP-40, 0.5%

751 sodium deoxycholate, 0.1% SDS). Chromatin was sheared using a Covaris sonicator for 5 min
752 (30 sec on/30 sec off) with amplitude=40. Correct chromatin fragmentation was confirmed using
753 1% agarose gels. 50 µg of chromatin was subjected to IP overnight at 4°C with 4 µg of anti-Oct1
754 (Novus Biological), Oct4 (Santa Cruz) or H3K27me3 (Milipore) antibodies. As a control, we used
755 5 µg of sheared, non-precipitated input chromatin. Samples were incubated with protein G
756 magnetic beads (ThermoFisher) for 5 hr and washed in Low Salt buffer (20 mM Tris-Cl pH 8.0,
757 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), High Salt buffer (identical but with 500
758 mM NaCl), LiCl buffer, and Tris-EDTA pH 8.0 plus 1 mM EDTA (TE buffer). Washes were
759 performed at 4°C for 10 min with rotation. For re-ChIP, 2% fragmented chromatin was saved as
760 input and the rest used for IP with the primary antibody (either Utx or Oct1) at 4°C overnight on a
761 rotator. Samples were then incubated with magnetic beads for 5h at 4°C. The beads were washed
762 for 10 min with Low Salt buffer, High Salt buffer, Low Salt buffer, LiCl buffer, and TE buffer
763 sequentially at 4°C. Chromatin was eluted with 300 μL IP Elution buffer at RT, then diluted 10-
764 fold in RIPA buffer. Diluted samples were then subjected to a second IP with 4 μg of the secondary
765 antibody (Oct1 or Oct4) at 4°C overnight, and incubated with magnetic beads for 5 hr at 4°C. The
766 beads were washed again as described above, then eluted with 300 μL IP Elution Buffer at RT.
767 Re-ChIP samples, together with the 2% input, were incubated at 65°C overnight to reverse
768 crosslinking. DNA was purified using phenol-chloroform-isoamyl alcohol extraction followed by
769 PCR clean up. qPCR primers can be found in Table S7 and were confirmed to generate a single
770 product of the correct size. The results were reported as qPCR values normalized to input
771 chromatin (gDNA) and non-specific region and presented as fold enrichment.
772
773 **ChIP-seq analysis.** After chromatin was precipitated as described above, and libraries were
774 sequenced using Illumina NovaSeq. Between 22 and 26 million paired-end Illumina sequencing
775 reads were aligned to the mouse *Mm10* reference genome using Novocraft novoalign v3.8.2,
776 allowing for one random alignment of multi-mapping reads, and providing the adapter sequences
777 for automatic trimming concordant with alignment. ChIP was analyzed using the
778 MultiRepMacsChIPSeq pipeline v12.2, using options "--pe --optdist 10000 --dupfrac 0.15 --mapq
779 10 --cutoff 2 --tdep 10 --peaksize 500 --peakgap 100 --species mouse --chrskip 'chrMIPhiX' --
780 blacklist mm10.blacklist.bed".
781
782 **Hi-ChIP and analysis.** The three-dimensional (3D) genome structure was analyzed using Hi-
783 ChIP, which combines chromosome conformation capture with chromatin immunoprecipitation.

784 Hi-ChIP was performed as previously described (Mumbach et al., 2016) using H3K27Ac
785 antibodies (Abcam, ab4729) and two biological replicates for each condition. The restriction
786 enzyme *Dpn*II was used instead of *Mbo*I. Crosslinked chromatin was sonicated using an EpiShear
787 probe-in sonicator (Active Motif) with three cycles of 30 sec at an amplitude of 40% with 30
788 seconds rest between cycles. Hi-ChIP libraries were sequenced using a NovaSeq 6000 as paired-
789 end 50 basepair reads to average depth of 300–400 million read-pairs per lane (100-200 million
790 reads per sample). Reads were aligned to the *Mm10* reference genome using HiC-Pro pipeline
791 (Servant et al., 2015) to extract informative unique paired-end tags (PETs). Hichipper (Lareau
792 and Aryee, 2018) was used to perform restriction site bias-aware modeling of the output from
793 HiC-Pro and to call interaction loops. These loops were filtered where loops with >1 reads in both
794 biological replicates were kept and used for downstream analyses. Chromatin interactions that
795 significantly change with cellular differentiation were identified using DESeq2 with an adjusted *P*-
796 value cutoff of 0.1. Loop anchors were assigned to the nearest gene using GREAT with a cutoff
797 of 100kb.
798
799 **Immunoprecipitation.** Cells were lysed with Cell Lysis Buffer (Life Technologies) in the presence
800 of protease inhibitors (EDTA-free tablet, Roche). IP was performed using 500 µg of extract.
801 Extracts were incubated with 4 µg of anti-Utx (Cell Signaling, D3Q1I) or Oct1 (Novus Biologicals,
802 NBP2-21584) antibodies, or rabbit IgG control overnight at 4°C. Protein-antibody complexes were
803 precipitated with protein-G magnetic beads (Thremo Fisher) for 3 hr at 4°C with rotation and
804 washed 3 times with Low Salt buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1%
805 SDS, 1% Triton X-100) plus protease inhibitors. Precipitated proteins were analyzed by
806 immunoblot.
807
808 **Luciferase reporter assay.** Single or combination mutations were introduced into the Oct1 and
809 Smad consensus binding sites in the following mouse *Myog* regulatory element (*Mm10*
810 chr1:134,285,666-134,285,750). IDT g-blocks® were synthesized to contain WT sequences or
811 single or combined mutations in the Oct1 or Smad binding sites, fused upstream of the CMV basal
812 promoter (-216-13 relative to TSS). G-blocks were inserted using sequence- and ligase-
813 independent cloning (Li and Elledge, 2012) upstream of the coding sequence for a secreted nano-
814 luciferase following digestion of vector pNL2.3 (Promega) using *Eco*RV and *Hind*III. The veracity
815 of the cloned inserts was confirmed by Sanger sequencing. 200 ng of reporter plasmid were co-
816 transfected into Oct1-deficient MEFs (Shakya et al., 2009) in DMEM media lacking phenol red

817 (ThermoFisher) together with 400 ng MMP9-mCherry (Wider and Picard, 2017) in 1000 ng total
818 transfected DNA. Where indicated, 400 ng pBabePuro-Oct1 was included in the transfection mix.
819 pUC18 plasmid comprised the balance of the transfected DNA. Where indicated, transfected cells
820 were provided with 5 ng recombinant mouse TGFb1 protein (R&D Systems). mCherry
821 fluorescence was determined first by exciting at 570 nm and measuring emission at 610 nm with
822 a 100 msec time delay using an Envision Xcite Multilabel Plate Reader. Luminescence was
823 measured using Nano-Glo Luciferase (Promega) and a Modulus luminescence plate reader.

824 **Acknowledgements**

825 We thank G. Kardon and C. Kikani for critical reading of the manuscript. We thank B. Dalley and
826 O. Allen at the HCI High-Throughput Genomics facility and T. Parnell and B. Lohman at the HCI
827 Bioinformatic Analysis Shared Resource for assistance with ChIP-seq and scRNA-seq. We thank
828 Olivier Pourquié from the Harvard Medical School for the assistance with the mesodermal
829 differentiation protocols. We thank D. Picard and D. Wider for the MMP9-mCherry construct. This
830 work was supported by a grant from the National Institutes of Health/National Institute of General
831 Medical Sciences (R01GM122778) to DT. The funders had no role in study design, data collection
832 and analysis, decision to publish, or preparation of the manuscript.

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

837

838 **Conflict of interest**

839 The authors declare that they have no conflicts of interest

840 **REFERENCES**

841

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

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

906 doi:10.1038/s41586-020-2493-4
907 Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI. 1995. Embryonic stem cells express
908 neuronal properties in vitro. *Dev Biol* **168**:342–357. doi:10.1006/dbio.1995.1085
909 Beddington RSP, Robertson EJ. 1989. An assessment of the developmental potential of
910 embryonic stem cells in the midgestation mouse embryo. *Development* **105**:733–737.
911 Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M,
912 Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. 2006. A bivalent
913 chromatin structure marks key developmental genes in embryonic stem cells. *Cell*
914 **125**:315–26. doi:10.1016/j.cell.2006.02.041
915 Chal J, Oginuma M, Tanoury Z Al, Gobert B, Sumara O, Hick A, Bousson F, Zidouni Y, Mursch
916 C, Moncuquet P, Tassy O, Vincent S, Miyanari A, Bera A, Garnier J, Guevara G, Hestin M,
917 Kennedy L, Hayashi S, Drayton B, Cherrier T, Gayraud-morel B, Gussoni E, Relaix F,
918 Tajbakhsh S, Pourquié O. 2015. Differentiation of pluripotent stem cells to muscle fiber to
919 model Duchenne muscular dystrophy. *Nat Biotechnol* **33**:962–9. doi:10.1038/nbt.3297
920 Cloos PAC, Christensen J, Agger K, Helin K. 2008. Erasing the methyl mark: Histone
921 demethylases at the center of cellular differentiation and disease. *Genes Dev*.
922 doi:10.1101/gad.1652908
923 Collinson A, Collier AJ, Morgan NP, Sienerth AR, Chandra T, Andrews S, Rugg-Gunn PJ. 2016.
924 Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and
925 Differentiation Defects in Human Embryonic Stem Cells. *Cell Rep*.
926 doi:10.1016/j.celrep.2016.11.032
927 Conlon FL, Lyons KM, Takaesu N, Barth KS, Kispert A, Herrmann B, Robertson EJ. 1994. A
928 primary requirement for nodal in the formation and maintenance of the primitive streak in
929 the mouse. *Trends Genet* **10**:308–308. doi:10.1016/0168-9525(94)90026-4
930 Dell'Orso S, Juan AH, Ko KD, Naz F, Perovanovic J, Gutierrez-Cruz G, Feng X, Sartorelli V.
931 2019. Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and
932 regenerative conditions. *Development* **146**. doi:10.1242/DEV.174177
933 DeVeale B, Brokhman I, Mohseni P, Babak T, Yoon C, Lin A, Onishi K, Tomilin A, Pevny L,
934 Zandstra PW, Nagy A, van der Kooy D. 2013. Oct4 is required ~E7.5 for proliferation in the
935 primitive streak. *PLoS Genet* **9**. doi:10.1371/JOURNAL.PGEN.1003957
936 Dhar SS, Lee SH, Chen K, Zhu G, Oh WK, Allton K, Gafni O, Kim YZ, Tomoiga AS, Barton MC,
937 Hanna JH, Wang Z, Li W, Lee MG. 2016a. An essential role for UTX in resolution and
938 activation of bivalent promoters. *Nucleic Acids Res* **44**:3659–3674.

939 doi:10.1093/nar/gkv1516
940 Dhar SS, Lee SH, Chen K, Zhu G, Oh WK, Allton K, Gafni O, Kim YZ, Tomoiga AS, Barton MC,
941 Hanna JH, Wang Z, Li W, Lee MG. 2016b. An essential role for UTX in resolution and
942 activation of bivalent promoters. *Nucleic Acids Res* **44**:3659–3674.
943 doi:10.1093/nar/gkv1516
944 Diaz-Cuadros M, Wagner DE, Budjan C, Hubaud A, Tarazona OA, Donelly S, Michaut A, Al
945 Tanoury Z, Yoshioka-Kobayashi K, Niino Y, Kageyama R, Miyawaki A, Touboul J,
946 Pourquié O. 2020. In vitro characterization of the human segmentation clock. *Nature*
947 **580**:113–118. doi:10.1038/s41586-019-1885-9
948 Gnocchi VF, White RB, Ono Y, Ellis JA, Zammit PS. 2009. Further characterisation of the
949 molecular signature of quiescent and activated mouse muscle satellite cells. *PLoS One*
950 **4**:e5205. doi:10.1371/journal.pone.0005205
951 Guibentif C, Griffiths JA, Imaz-Rosshandler I, Ghazanfar S, Nichols J, Wilson V, Göttgens B,
952 Marioni JC. 2021. Diverse Routes toward Early Somites in the Mouse Embryo. *Dev Cell*
953 **56**:141-153.e6. doi:10.1016/j.devcel.2020.11.013
954 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass
955 CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-
956 Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**:576–589.
957 doi:10.1016/j.molcel.2010.05.004
958 Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. 2013.
959 Super-enhancers in the control of cell identity and disease. *Cell* **155**:934–47.
960 doi:10.1016/j.cell.2013.09.053
961 Kang J, Gemberling M, Nakamura M, Whitby FG, Handa H, Fairbrother WG, Tantin D. 2009a. A
962 general mechanism for transcription regulation by Octl and Oct4 in response to genotoxic
963 and oxidative stress. *Genes Dev* **23**:208–222. doi:10.1101/gad.1750709
964 Kang J, Gemberling M, Nakamura M, Whitby FG, Handa H, Fairbrother WG, Tantin D. 2009b. A
965 general mechanism for transcription regulation by Oct1 and Oct4 in response to genotoxic
966 and oxidative stress. *Genes Dev* **23**:208–222. doi:10.1101/gad.1750709
967 Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum
968 C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein
969 BE. 2008. Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of
970 Bivalent Domains. *PLoS Genet* **4**:e1000242. doi:10.1371/journal.pgen.1000242
971 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

Lareau CA, Aryee MJ. 2018. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat Methods* **15**:155–156. doi:10.1038/NMETH.4583

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

Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY. 2016. HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*. doi:10.1038/nmeth.3999

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.

1005 doi:10.1016/j.devcel.2017.02.001
1006 Palmieri SL, Peter W, Hess H, Schöler HR. 1994. Oct-4 transcription factor is differentially
1007 expressed in the mouse embryo during establishment of the firsttwo extraembryonic cell
1008 lineages involved in implantation. *Dev Biol* **166**:259–267. doi:10.1006/dbio.1994.1312
1009 Perovanovic J, DellOrso S, Gnochi VF, Jaiswal JK, Sartorelli V, Vigouroux C, Mamchaoui K,
1010 Mouly V, Bonne G, Hoffman EP. 2016. Laminopathies disrupt epigenomic developmental
1011 programs and cell fate. *Sci Transl Med* **8**:335ra58-335ra58.
1012 doi:10.1126/scitranslmed.aad4991
1013 Schlesinger S, Meshorer E. 2019. Open Chromatin, Epigenetic Plasticity, and Nuclear
1014 Organization in Pluripotency. *Dev Cell*. doi:10.1016/j.devcel.2019.01.003
1015 Sebastiano V, Dalvai M, Gentile L, Schubart K, Sutter J, Wu GM, Tapia N, Esch D, Ju JY,
1016 Hübner K, Bravo MJA, Schöler HR, Cavaleri F, Matthias P. 2010. Oct1 regulates
1017 trophoblast development during early mouse embryogenesis. *Development* **137**:3551–
1018 3560. doi:10.1242/dev.047027
1019 Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E.
1020 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*
1021 **16**. doi:10.1186/S13059-015-0831-X
1022 Seydoux G, Braun RE. 2006. Pathway to Totipotency: Lessons from Germ Cells. *Cell*.
1023 doi:10.1016/j.cell.2006.11.016
1024 Shakya A, Callister C, Goren A, Yosef N, Garg N, Khoddami V, Nix D, Regev A, Tantin D. 2015.
1025 Pluripotency transcription factor Oct4 mediates stepwise nucleosome demethylation and
1026 depletion. *Mol Cell Biol* **35**:1014–25. doi:10.1128/MCB.01105-14
1027 Shakya A, Cooksey R, Cox JE, Wang V, McClain DA, Tantin D. 2009. Oct1 loss of function
1028 induces a coordinate metabolic shift that opposes tumorigenicity. *Nat Cell Biol* **11**:320–327.
1029 doi:10.1038/ncb1840
1030 Shakya A, Kang J, Chumley J, Williams MA, Tantin D. 2011. Oct1 is a switchable, bipotential
1031 stabilizer of repressed and inducible transcriptional states. *J Biol Chem* **286**:450–9.
1032 doi:10.1074/jbc.M110.174045
1033 Shen Z, Kang J, Shakya A, Tabaka M, Jarboe EA, Regev A, Tantin D. 2017. Enforcement of
1034 developmental lineage specificity by transcription factor Oct1. *Elife* **6**.
1035 doi:10.7554/eLife.20937
1036 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M,
1037 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

Tantin D. 2013. Oct transcription factors in development and stem cells: insights and mechanisms. *Development* **140**:2857–2866. doi:10.1242/dev.095927

Tantin D, Schild-Poulter C, Wang V, Haché RJG, Sharp PA. 2005. The Octamer Binding Transcription Factor Oct-1 Is a Stress Sensor. *Cancer Res* **65**:10750–10758. doi:10.1158/0008-5472.CAN-05-2399

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381–386. doi:10.1038/nbt.2859

Wang C, Lee JE, Cho YW, Xiao Y, Jin Q, Liu C, Ge K. 2012. UTX regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity. *Proc Natl Acad Sci U S A* **109**:15324–15329. doi:10.1073/pnas.1204166109

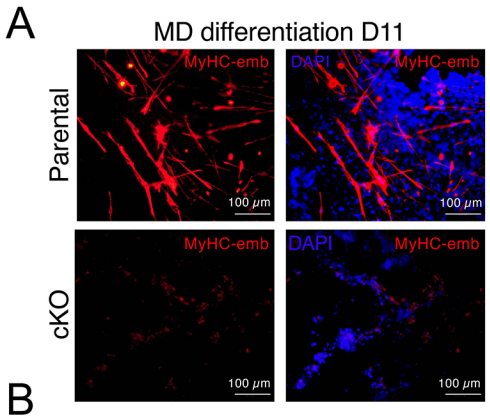
Wang VEH, Schmidt T, Chen J, Sharp PA, Tantin D. 2004. Embryonic lethality, decreased erythropoiesis, and defective octamer-dependent promoter activation in Oct-1-deficient mice. *Mol Cell Biol* **24**:1022–32. doi:10.1128/MCB.24.3.1022-1032.2004

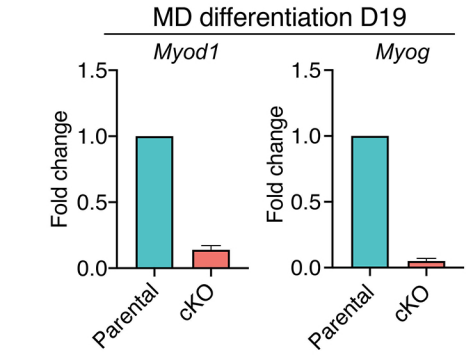
Wider D, Picard D. 2017. Secreted dual reporter assay with Gaussia luciferase and the red fluorescent protein mCherry. *PLoS One* **12**:e0189403. doi:10.1371/journal.pone.0189403

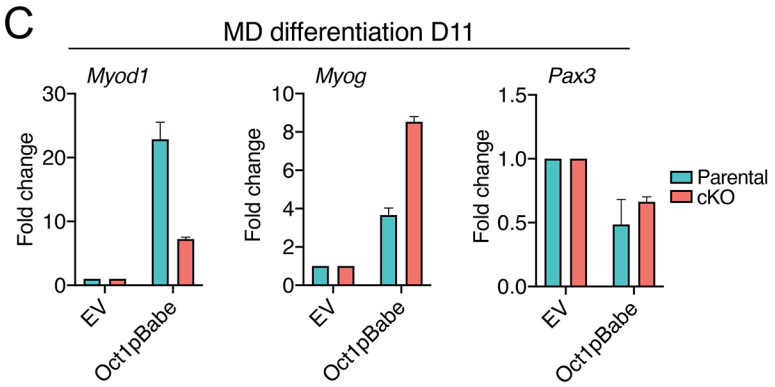
Zhang Y, Lahmann I, Baum K, Shimojo H, Mourikis P, Wolf J, Kageyama R, Birchmeier C. 2021. Oscillations of Delta-like1 regulate the balance between differentiation and maintenance of muscle stem cells. *Nat Commun* **12**. doi:10.1038/s41467-021-21631-4

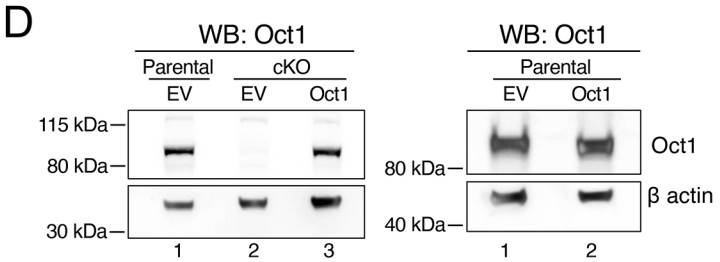
Zhou X, Sasaki H, Lowe L, Hogan BLM, Kuehn MR. 1993. Nodal is a novel TGF-β-like gene expressed in the mouse node during gastrulation. *Nature* **361**:543–547. doi:10.1038/361543a0

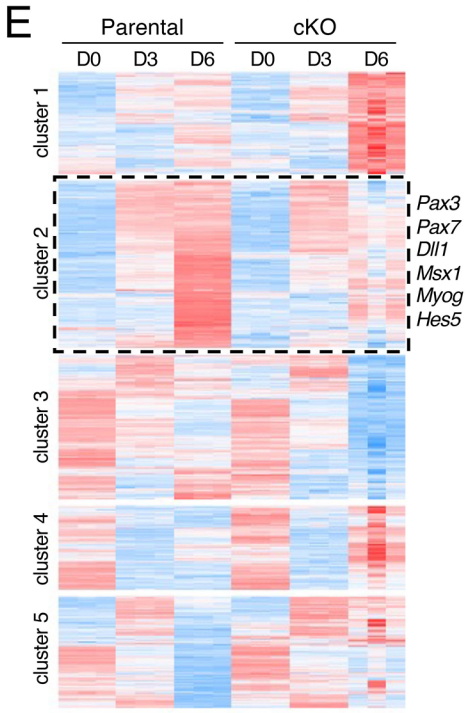
Perovanovic *et al.* Figure 1











F

Cluster 1 Enrichment

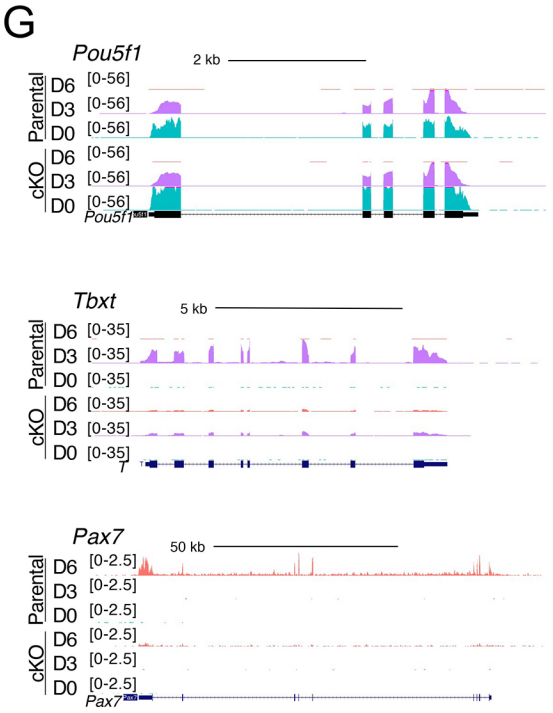
Jensen Tissue	Adj. *p*-value
Germ cell	2.8E-6
Primordium	1.5E-5

Cluster 2 Enrichment

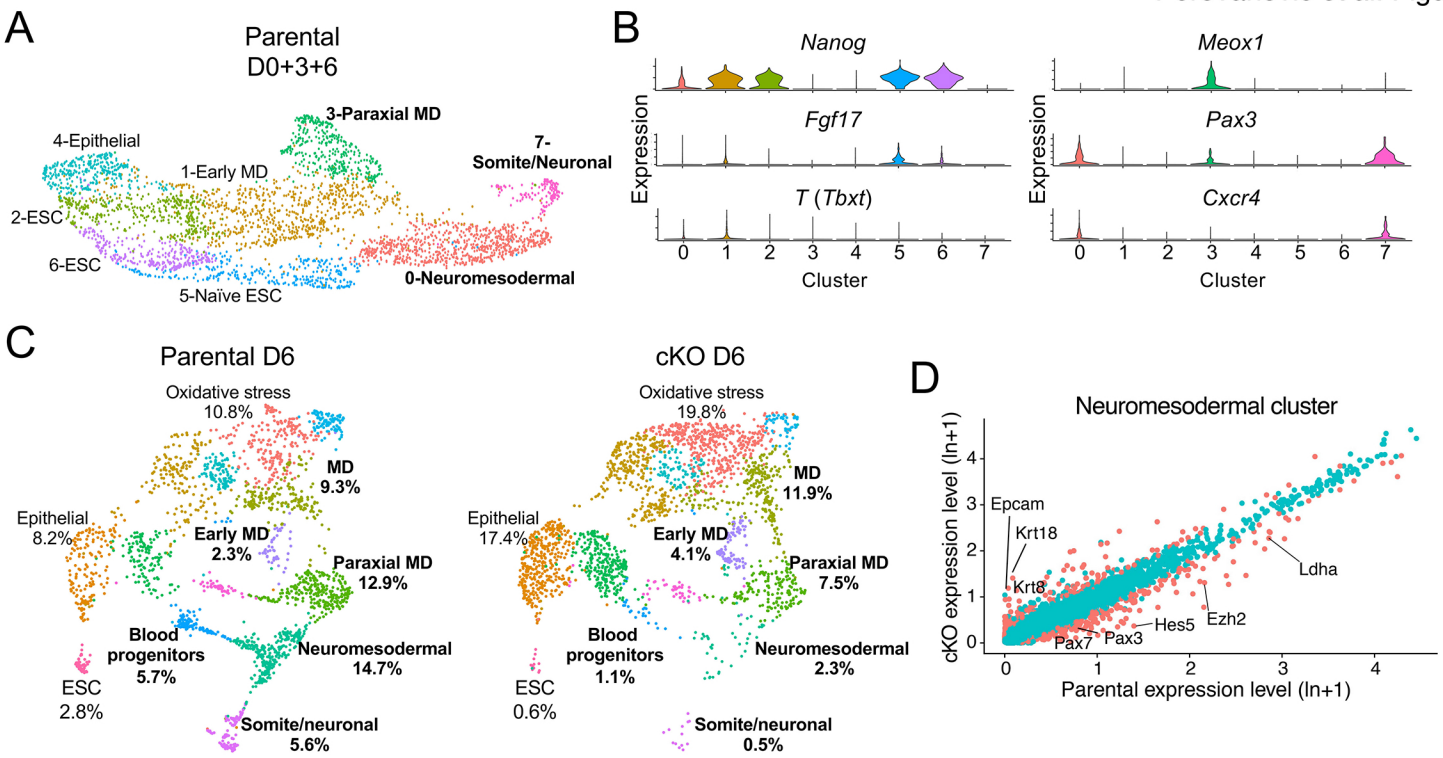
Jensen Tissue	Adj. *p*-value
Neural Tube	2.8E-6
Somite	3.8E-5

Cluster 2 Regulation

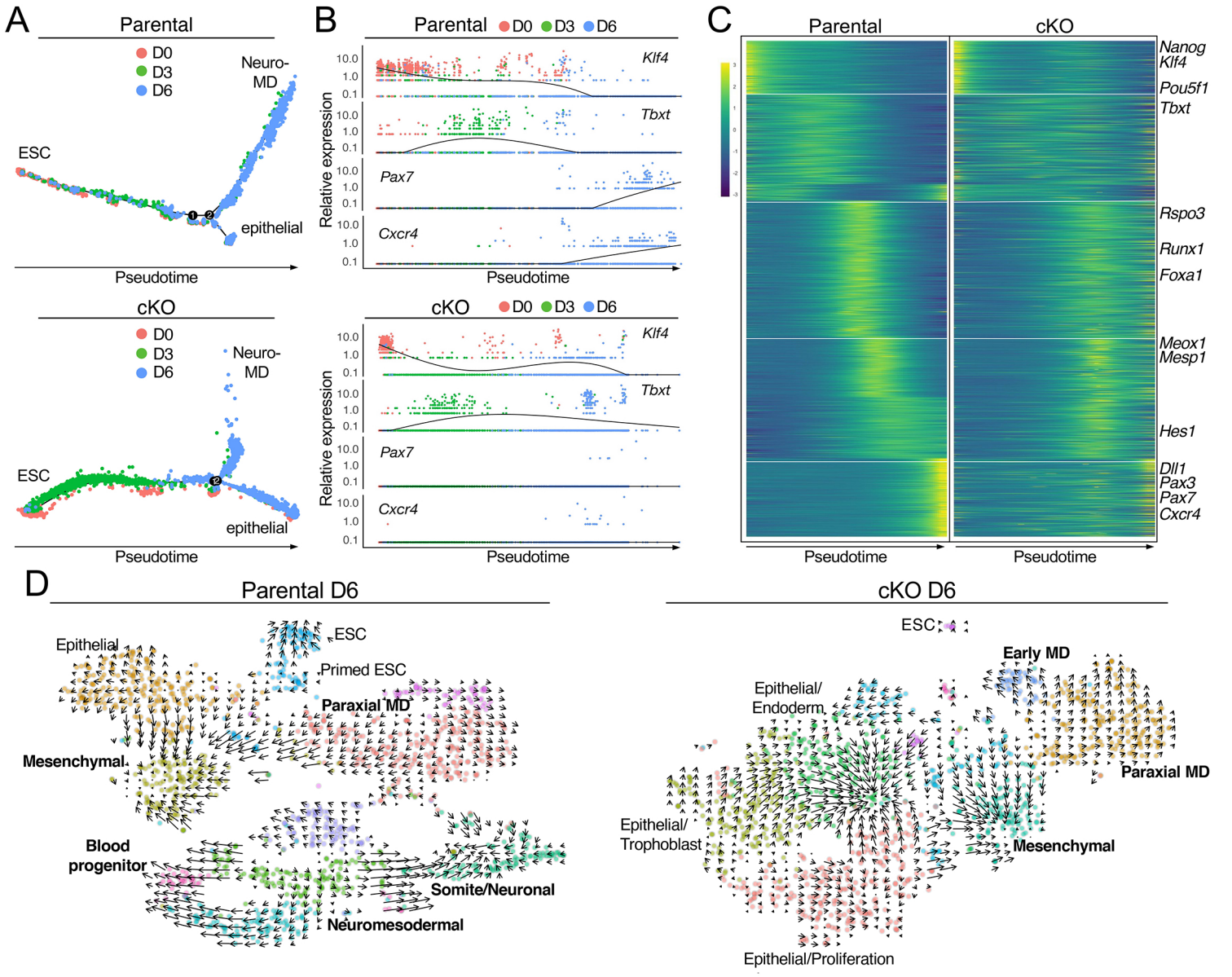
ChEA 2016	Adj. *p*-value
SUZ12	1.836E-21
MTF2	7.127E-21
JARID2	2.807E-20
EZH2	7.379E-20



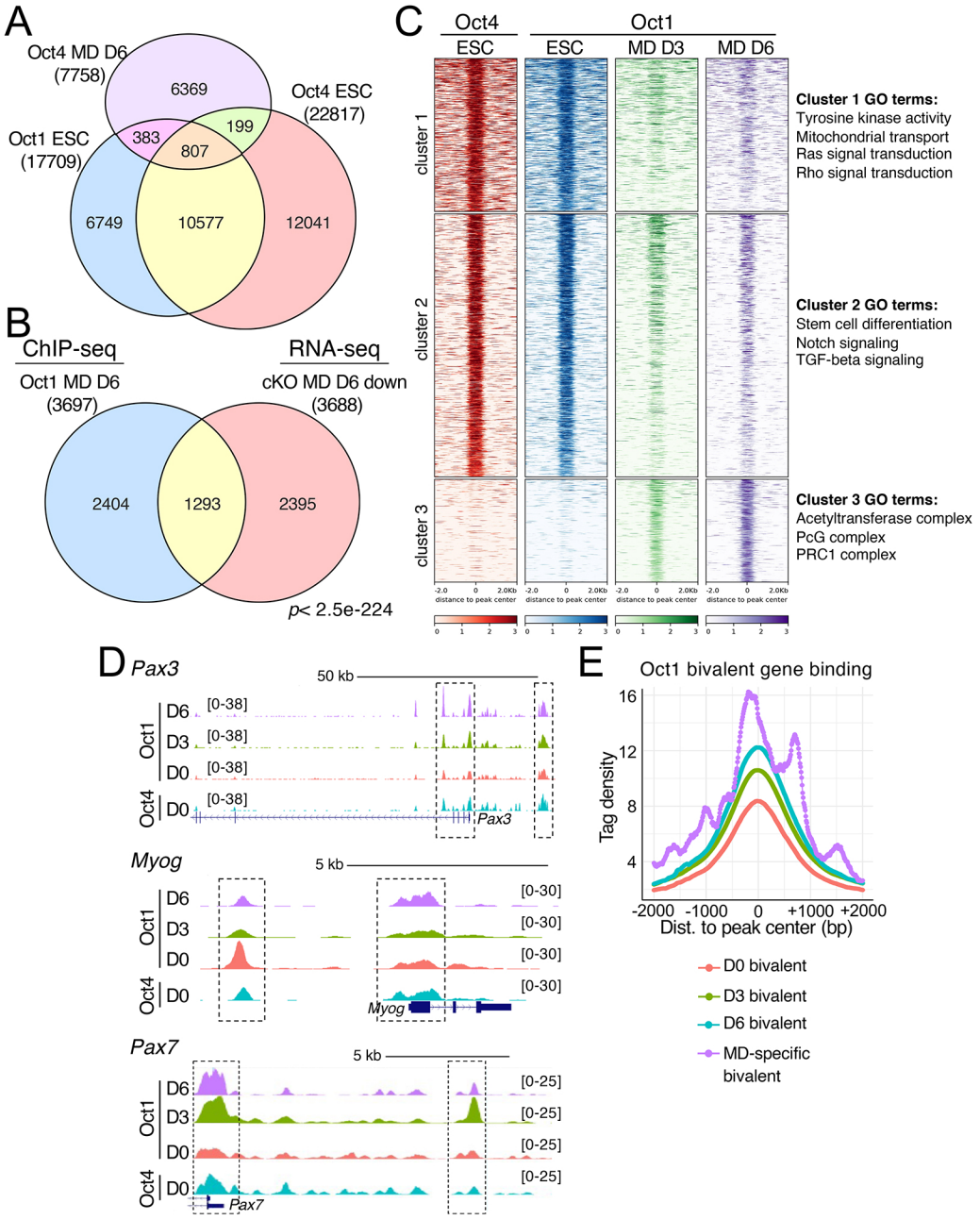
Perovanovic *et al.* Figure 2



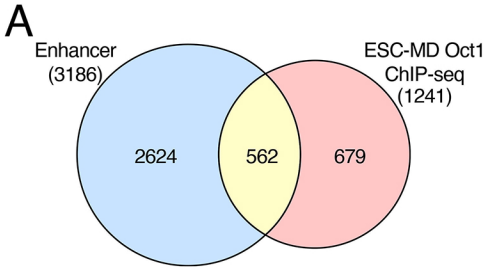
Perovanovic *et al.* Figure 3

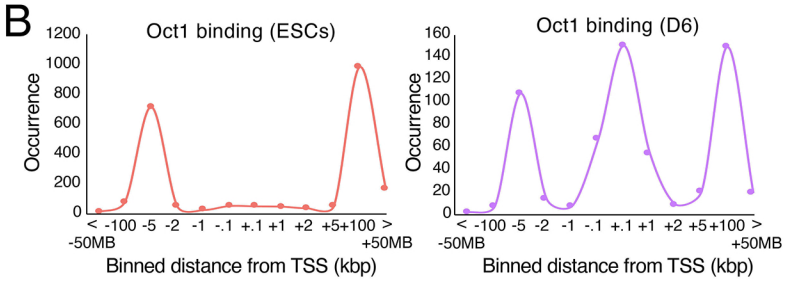


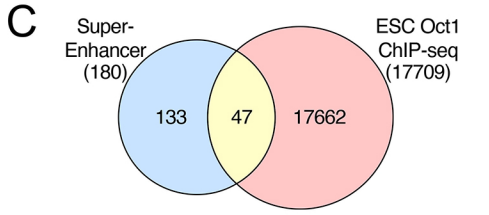
Perovanovic *et al.* Figure 4



Perovanovic *et al.* Figure 5

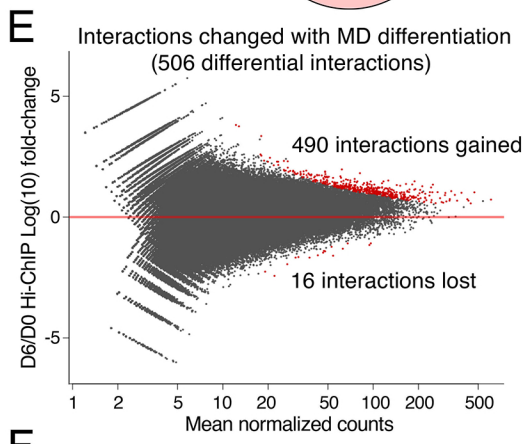


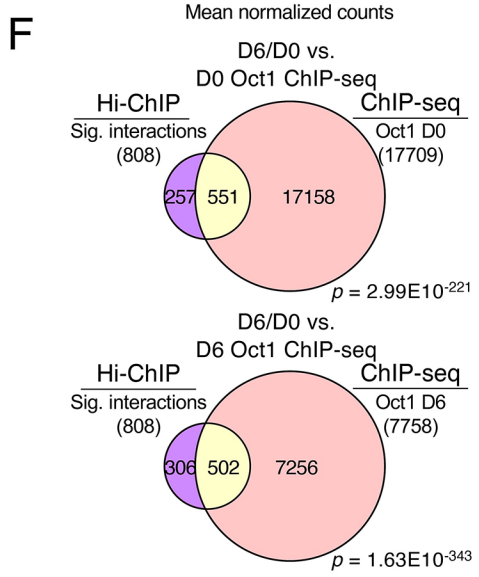


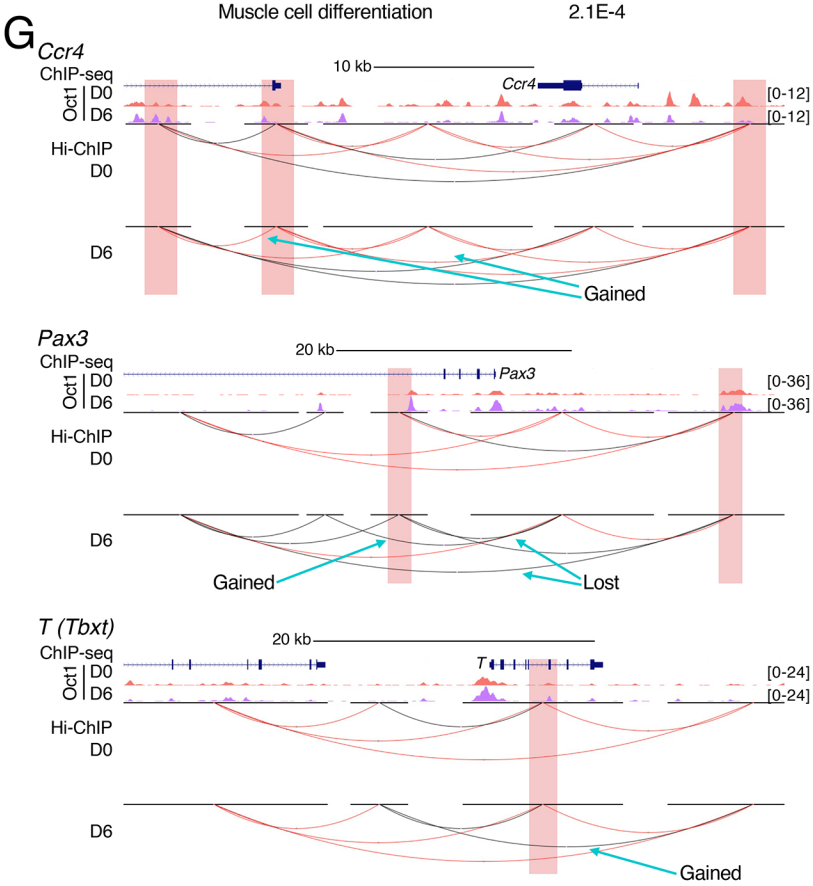


D

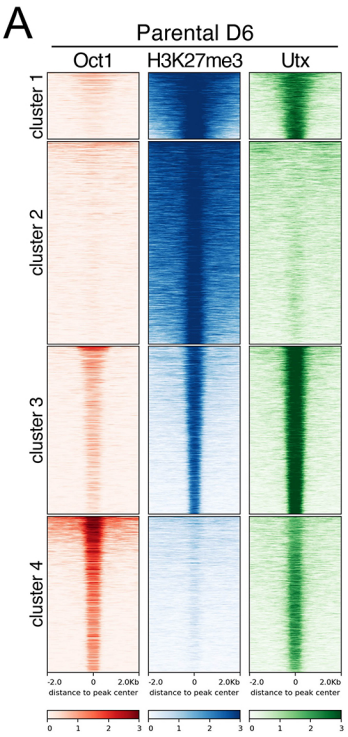
Terms associated with Oct1 SE genes	Binom FDR Q
Embryonic organ morphogenesis	9.4E-6
Embryonic morphogenesis	8.3E-6
Cardiomyocyte differentiation	9.0E-6
Bundle of His development	8.3E-5
His-Purkinje system development	9.9E-5
Cardiac muscle tissue development	1.3E-4
Muscle cell differentiation	2.1E-4







Perovanovic *et al.* Figure 6



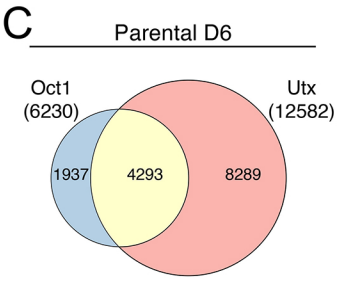
B

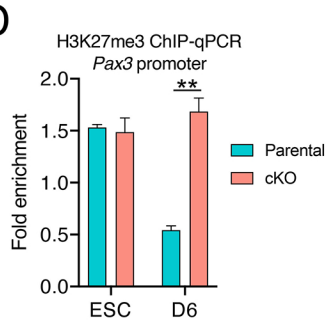
Cluster 3 Enrichment

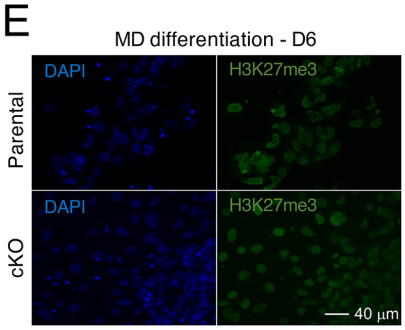
Human Phenotype Ontology	Binome FDR
Abnormality of the musculature	2.00E-46
Abnormality of muscle physiology	1.20E-44
Abnormality of skeleton morphology	7.51E-41

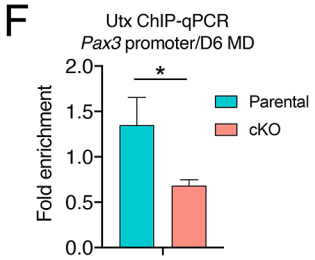
Cluster 4 Enrichment

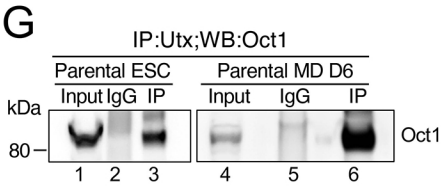
Human Phenotype Ontology	Binome FDR
Abnormality of the musculature	1.99E-39
Global developmental delay	1.72E-36
Neurodevelopmental delay	2.82E-35
Abnormality of metabolism	5.28E-35

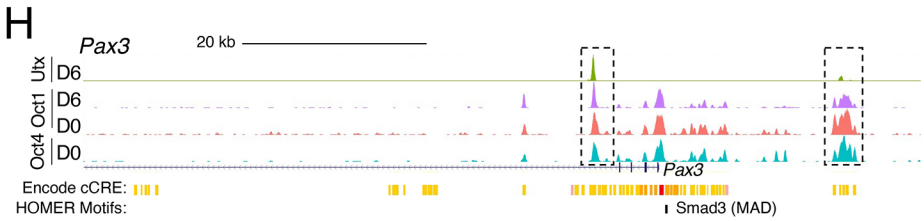
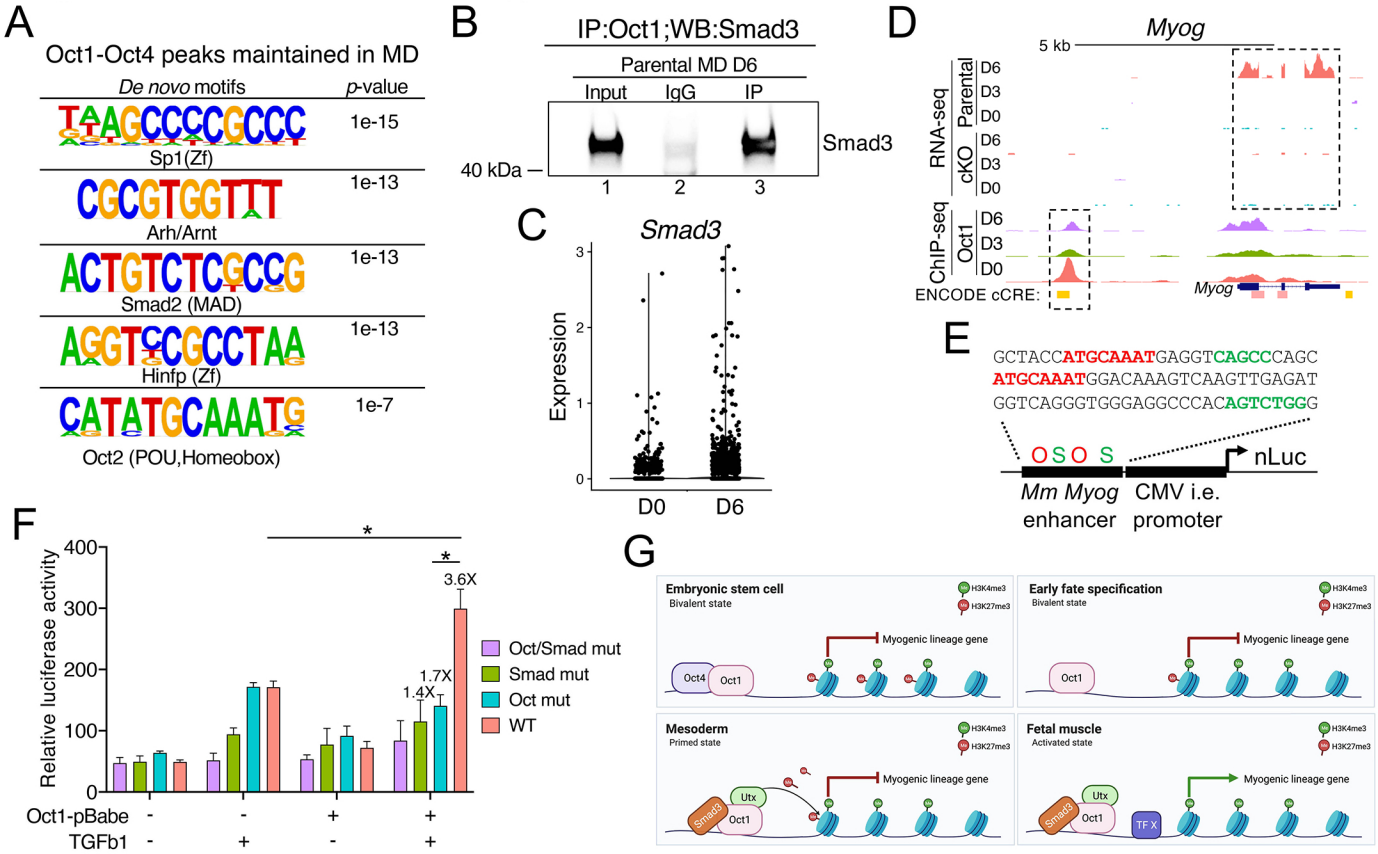












Perovanovic *et al.* Figure 6