Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2022 November 09.

Published in final edited form as:

Cell Rep. 2022 August 16; 40(7): 111219. doi:10.1016/j.celrep.2022.111219.

Transcriptomics, regulatory syntax, and enhancer identification in mesoderm-induced ESCs at single-cell resolution

Mamduh Khateb[1], **Jelena Perovanovic**[1,9], **Kyung Dae Ko**[1,9], **Kan Jiang**[2,9], **Xuesong Feng**[1], **Natalia Acevedo-Luna**[1], **Jérome Chal**[3,4,5], **Veronica Ciuffoli**[1], **Pavol Genzor**[6], **James Simone**[7], **Astrid D. Haase**[6], **Olivier Pourquié**[3,4,5], **Stefania Dell'Orso**[8,*], **Vittorio Sartorelli**[1,10,*]

[1]Laboratory of Muscle Stem Cells and Gene Regulation, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), NIH, Bethesda, MD, USA

[2]Biodata Mining and Discovery Section, NIAMS, NIH, Bethesda, MD, USA

[3]Department of Genetics, Harvard Medical School, Boston, MA, USA

[4]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA

[5]Harvard Stem Cell Institute, Boston, MA, USA

[6]National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD, USA

[7]FlowCytometry Section, NIAMS, NIH, Bethesda, MD, USA

[8]Genomic Technology Section, NIAMS, Bethesda, MD, USA

[9]These authors contributed equally

[10]Lead contact

SUMMARY

Embryonic stem cells (ESCs) can adopt lineage-specific gene-expression programs by stepwise exposure to defined factors, resulting in the generation of functional cell types. Bulk and single-cell-based assays were employed to catalog gene expression, histone modifications, chromatin conformation, and accessibility transitions in ESC populations and individual cells acquiring a presomitic mesoderm fate and undergoing further specification toward myogenic and neurogenic lineages. These assays identified *cis*-regulatory regions and transcription factors presiding over

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

*Correspondence: stefania.dellorso@nih.gov (S.D.), vittorio.sartorelli@nih.gov (V.S.).

AUTHOR CONTRIBUTIONS

M.K. performed most of the experiments. J.P. conceived the study, established initial ESC culture conditions, and performed and analyzed ATAC-seq and bulk RNA-seq experiments. K.D.K. and K.J. analyzed RNA-seq, ATAC-seq, ChIP-seq, scRNA-seq, scATAC-seq, and multiomics data. X.F. isolated somites. N.A.-L. analyzed Hi-C data. V.C. helped with libraries preparation. P.G. and A.D.H. provided advice on CRISPR-Cas9 experiments. J.S. assisted with FACS. J.C. and O.P. provided assistance for optimization of ESC differentiation protocols. S.D.O. provided advice for the genomic studies, analyzed data, and supervised the project. V.S. conceived and designed the study, analyzed data, supervised the project, and wrote the manuscript with input from the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111219.

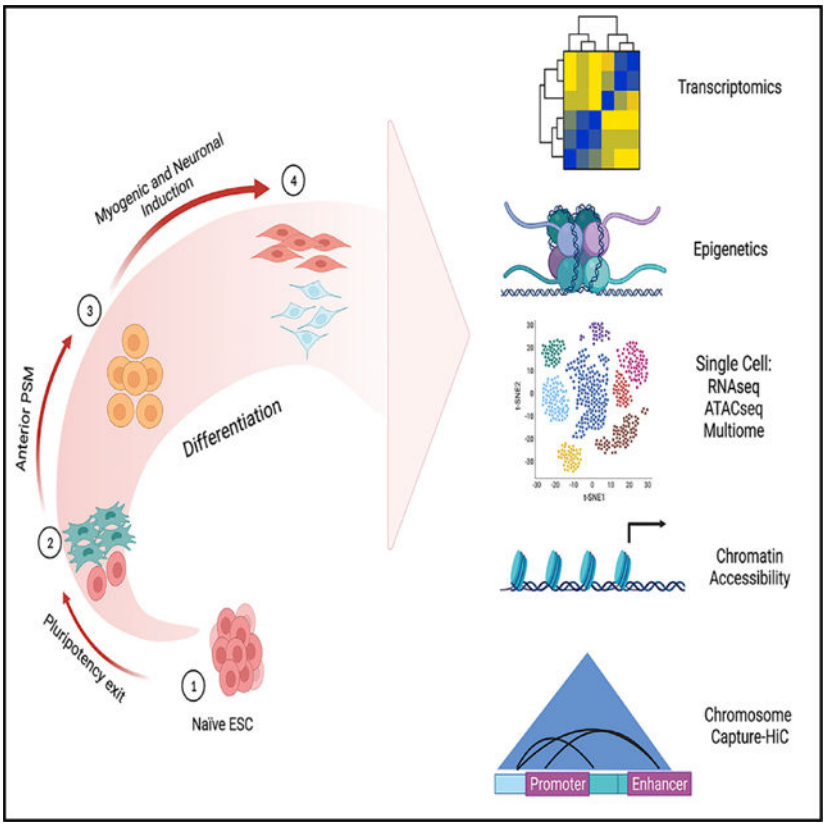
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

gene-expression programs occurring at defined ESC transitions and revealed the presence of heterogeneous cell populations within discrete ESC developmental stages. The datasets were employed to identify previously unappreciated genomic elements directing the initial activation of Pax7 and myogenic and neurogenic gene-expression programs. This study provides a resource for the discovery of genomic and transcriptional features of pluripotent, mesoderm-induced ESCs and ESC-derived cell lineages.

In brief

Khateb et al. provide bulk and single-cell transcriptional, epigenetic, and chromatin accessibility profiling of ESCs induced to acquire presomitic mesoderm, myogenic, and neurogenic fates and identify a Pax7 enhancer directing transcription in myogenic and neurogenic precursors.

Graphical Abstract



INTRODUCTION

Embryonic stem cells (ESCs) self-renew indefinitely, retain pluripotency, and differentiate into all adult cell types without changes in their genetic information (Martello and Smith, 2014). DNA and histone chemical modifications as well as chromatin accessibility and architecture collectively ensure genome stability, correct propagation of genetic information, and proper interpretation of the genome (Allis and Jenuwein, 2016). Enhancers play a key role in these processes by orchestrating cell-type-specific gene-expression programs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Visel et al., 2009; Catarino and Stark, 2018). Muscle stem cells (MuSCs), which are required for muscle growth and regeneration (Brack and Rando, 2012; Yin et al., 2013; Tierney and Sacco, 2016; Evano and Tajbakhsh, 2018; Fuchs and Blau, 2020; Relaix et al., 2021), can be generated from mouse or human pluripotent stem cells by either treatment with defined molecules or forced expression of the transcription factor Pax7 (Darabi et al., 2012; Chal et al., 2015, 2018; Al Tanoury et al., 2020; Xi et al., 2020; Magli and Perlingeiro, 2017). Using a protocol mimicking key signaling events occurring in the mouse embryo, we aimed at providing a resource that can be utilized to analyze transcriptional, epigenetics, chromatin accessibility, and conformation events accompanying exit from ESC pluripotency, acquisition of anterior presomitic mesoderm (aPSM) fate, and further myogenic and neurogenic differentiation. Integration of the data we generated uncovered chromatin dynamics underlying activation of cell-fate programs in ESC-derived precursors confirmed to occur in somites of mouse developing embryos. Bulk and single-cell analysis revealed distinct molecular syntax utilized by regulatory regions to control expression of individual genes within a common gene-regulatory network. We leveraged these datasets and, by combining chromatin accessibility, *in situ* Hi-C chromatin conformation, genome editing, and reporter assays, have identified regulatory regions directing initial Pax7 expression and activation of the myogenic and neurogenic programs.

RESULTS

Instructing ESCs

Mouse ESCs in which the green fluorescent protein (GFP) coding region has been placed under the control of endogenous Pax3 regulatory sequences (ESCs Pax3-GFP) (Chal et al., 2015) were cultured in conditions maintaining pluripotency (feeder-free conditions, LIF+2i inhibitors, naive ESCs) (Ying et al., 2008). BMP4 is essential for embryogenesis, predominantly for mesoderm formation (Dale et al., 1992), and drives commitment to differentiation of human ESCs (Gunne-Braden et al., 2020). We induced initial ESC differentiation employing a medium supplemented with Bmp4 protein (10 ng/ml) and 1% knockout serum replacement (KSR) for 48 h (instructed ESCs) (see STAR Methods) (Figures 1A and S1A). RNA sequencing (RNA-seq) analysis (1.5-fold change, adjusted $p < 0.05$) revealed 2,022 upregulated and 1,288 downregulated genes in instructed ESCs (Figures 1B; Table S1). Gene Ontology of upregulated transcripts returned terms related to cholesterol biosynthesis, cell morphogenesis, steroid biosynthesis, actin cytoskeleton organization, and Rho GTPase cycle (Figure S1B; Table S1). Downregulated transcripts in instructed ESCs were enriched for terms related to lysosome, autophagy, and pluripotency (Figure S1B; Table S1). Transcription factors Otx2 and Pou3f1 (Oct6), fibroblast growth factors Fgf5 and Fgf15, DNA methyltransferases Dnmt3a and Dnmt3b, and secreted factor Wnt8a, which mark an intermediate state (formative pluripotency) of ESCs transitioning from naive to primed pluripotency *in vitro* and *in vivo* (Buecker et al., 2014; Boroviak et al., 2015; Acampora et al., 2016; Kinoshita et al., 2021; Wang et al., 2021; Pera and Rossant, 2021), were upregulated in instructed ESCs (Figures 1C, 1E, and 1F). Moreover, a core of transcripts enriched in human pluripotent stem cells skewed toward a naïve-to-primed intermediate state (Cornacchia et al., 2019) was also increased in instructed ESCs (Table S1, UP-Lipid_Signaling). Instead, pluripotency factors Nanog, Esrrb, Klf4, Prdm14,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Tbx3, and Zfp42, downregulated in ESCs with epiblast-like primed pluripotency (Buecker et al., 2014), were reduced (Figures 1D–1F). Altogether, these data indicate that Bmp4 instructs ESCs to downregulate pluripotency genes and to express transcripts associated with fate specification, including a selected gene-expression profile shared with ESC formative intermediate states.

Identification of the regulatory landscape in instructed ESCs

Chromatin accessibility to transcription factors and cofactors is essential for establishing and maintaining cellular identity (Klemm et al., 2019). Surveying chromatin accessibility of naive and instructed ESCs by assay for transposase-accessible chromatin (ATAC)-seq revealed that 68% (1,385/2,022) of genes upregulated in instructed ESCs gained increased chromatin accessibility at their transcription start site (TSS) (Figures 1G and 1H). Of the 1,288 downregulated genes, 37% (479/1,288) displayed reduced chromatin accessibility (Figures 1G and 1H). Cell-type- and cell-state-specific gene expression requires coordination between promoters and enhancers (Field and Adelman, 2020). To identify active enhancers, we conducted H3K4me1 and H3K27ac chromatin immunoprecipitation (ChIP)-seq (Rada-Iglesias et al., 2011) (Figures 1I and S1C) and integrated the resulting datasets with ATAC-seq. Of the 1,482 instructed ESCs-specific enhancer regions (defined as H3K4me1^{+}/H3K27ac^{+}/ATAC^{+}), 25% were associated with increased transcription of the nearest gene (Tables S1 and S2). Albeit observed in a different experimental setting, these results agree with a study reporting that only a subset of genomic regions marked by H3K4me1 and H3K27ac in ESCs displayed enhancer activity (Barakat et al., 2018). Enhancers were activated with different modalities. In naive ESCs, 40% (597/1,482) of accessible enhancers had low H3K4me1 and were ATAC^{+}, 26% (386/1,482) of poised enhancers were H3K4me1^{+}/ATAC^{+}, 25.5% (379/1,482) of unmarked enhancers had the lowest H3K4me1^{-} and were ATAC^{-}, and 8% (120/1,482) of *de novo* instructed ESCs enhancer*s* were H3K4me1^{+}/ATAC^{-} (Figures 1J and 1K; Table S2). All 1,482 regions acquired H3K27ac in instructed ESCs (Figure S1D). Accessible enhancers were associated with 38%, poised enhancers with 29%, unmarked enhancers with 25%, and *de novo* enhancers with 7% of genes activated in instructed ESCs (Figure 1L). Here, we provide selected examples of enhancer activation and deactivation modalities. Enhancer and promoter regions of upregulated genes Pou3f1 (Oct6) and Wnt8a gained chromatin accessibility and H3K27ac (Figure S1E). Intronic enhancer regions of upregulated gene Pbx1 and Pitx2 genes were already accessible in naive conditions and acquired H3K27ac only upon ESC instruction. Selected pluripotent genes were downregulated in instructed ESCs (Figure 1E). H3K27ac decreased, while chromatin accessibility was preserved, at enhancers of Esrrb gene. H3K27ac and chromatin accessibility were both reduced at upstream regulatory regions of Nanog gene. In another instance, pluripotent Tbx3 gene was downregulated, and enhancer and promoter regions showed reduced H3K27ac and decreased chromatin accessibility. Thus, concordantly transcribed genes can be differently regulated. While frequently consonant with gene expression, chromatin accessibility and H3K27ac can be temporally disjointed, with chromatin accessibility often preceding transcription.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Single-cell analysis of naive and instructed ESCs

Because of the stochastic nature of the pluripotency network, ESCs are transcriptionally heterogeneous even when exposed to a uniform culture environment (Torres-Padilla and Chambers, 2014). Establishing whether extracellular signals monotonously induce gene-expression changes across a whole cell population or elicit pliable transcriptional responses in individual cells requires single-cell approaches. Uniform manifold approximation and projection (UMAP) was employed to construct a high-dimensional graph representation of single-cell RNA (scRNA)-seq datasets obtained from naive and instructed ESCs. Integration of the scRNA-seq datasets resulted in the identification of spatially distinct clusters that could be assigned to either naive or instructed ESCs and placed on a pseudotemporal map (Figure 1M). The main naive ESCs cluster (Figure 1M, cluster 1) expressed high levels of pluripotency genes, oxidative phosphorylation, and glycolytic transcripts, in line with the observation that pluripotent ESCs rely on both metabolic pathways for energy and metabolites production (Zhou et al., 2012; Ryall et al., 2015) (Table S2). Autophagy maintains ESC stemness, and Ulk1 (Atg1) is essential for this process (Gong et al., 2018). Expression of autophagy genes Ulk1 and Gabarapl2 (Atg8) was enriched in naive ESC clusters (Figure S1F). Two smaller additional clusters observed in naive ESCs (Figure 1M, clusters 3 and 5) were also enriched for pluripotency and ribosomal and translational transcripts (Table S2). Instructed ESCs generated two clusters located at distinct pseudotemporal points. The main instructed ESC cluster (Figure 1M, cluster 2) occupied an intermediate pseudotemporal space between naive ESCs and an additional instructed ESC cluster (cluster 4) and contained cells enriched for transcription factors (TFs) Pou3f1 (Oct6), Sox4, Otx2, and Sall2 (Figures 1N and S1G). Pluripotency genes displayed a graded expression in the different clusters. While overall reduced compared with naive ESCs, Nanog and Sox2 expression was still maintained in cluster 2 but drastically reduced in cluster 4 (Figure 1N). In contrast, Tbx3 displayed a bimodal expression in naive cluster 1 and in instructed cluster 4 (Figure 1N). This observation is consistent with a dual Tbx3 role in maintaining pluripotency and directing a cell-fate decision toward mesoendoderm (Weidgang et al., 2013). Instructed ESC clusters hosted cells sharing ~28% of genes belonging to embryonic day 8.5 (E8.5) and E9.5 neuromesodermal progenitor signatures (Gouti et al., 2017) (Table S2). Gata TFs can bind condensed chromatin and function as pioneer factors (Cirillo et al., 2002) (Sharma et al., 2020) and are coexpressed with the bookmarking TF Hnf1, which is involved in the mitotic transmission of epigenetic information (Verdeguer et al., 2010). Expression of Gata4, Gata6, Foxa2, and Hnf1b, which are involved in mesoendoderm induction (Rojas et al., 2005; Wamaitha et al., 2015; Coffinier et al., 1999; Costello et al., 2015), were enriched in cluster 4 (Figures 1O and S1I). scATAC-seq revealed that DNA-binding motifs (DBMs) for Gata4, Gata6, Foxa2, and Hnf1b were overrepresented and footprinted. Gata4 and Hnf1b reciprocally occupied their TSSs (Figures 1P, S1H, and S1I). Thus, a single-cell approach documented transitional states occurring in ESCs exposed to Bmp4 and identified TFs involved in these transitions.

Single-cell analysis of Pax3-expressing cells

Instructed ESCs were switched to culture conditions favoring acquisition of aPSM fate (R-spondin3 and the Bmp inhibitor LDN193189, RDL; see STAR Methods) (Figure S2A) (Chal et al., 2015). After 96 h in aPSM-promoting medium, ~30%–40% of ESC Pax3-GFP

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

became GFP$^+$ and could be fluorescence-activated cell sorting (FACS) isolated (Figures 2A and 2B, aPSM cells). scRNA-seq analysis performed on FACS-isolated Pax3-GFP cells identified four cell clusters that were ordered on a pseudotime map (Figure 2C). Posterior PSM (pPSM) genes were not expressed (Figure 2D), while aPSM genes were highly expressed, in the main cell clusters (clusters 0 and 1) (Figure 2E). Integration of scRNA-seq and scATAC-seq data revealed that, while not transcribed, the chromatin of pPSM genes was accessible (Figure 2D). Since pPSM genes are expressed in ESCs cultured in aPSM-promoting medium for 48 h (Chal et al., 2015), chromatin accessibility at their promoter regions may indicate memory preservation of a previously active, now extinct, gene program. For instance, pPSM Wnt5a, Wnt5b, Hoxd11, and Hoxd13 genes were not expressed while their chromatin was accessible (Figure 2D). Fgf8, which activates Hoxd11 and Hoxd13 (Rodrigues et al., 2017), was hardly detected in aPSM cells (Figure S2B), potentially explaining the lack of Hoxd11-Hoxd13 activation despite their permissive chromatin. Two additional clusters were identified (Figure 2C). Cluster 3 was composed of cells expressing endothelial-related genes Flt1, Flt4, and Kdr (Figures 2F and S2C). Cluster 2 hosted cells enriched for glycolytic genes and neuronal markers Pax6, Sox2, Sox11, and Nestin (Figures 2G, 2H, and S2D). Pax6, Sox2, and Nestin are highly expressed in neural stem cell precursors (NSCs), with Pax6 and Sox2 controlling NSC identity and differentiation (Gotz et al., 1998; Gomez-Lopez et al., 2011). Glycolysis prevents NSC precocious differentiation (Lange et al., 2016), raising the possibility that activated glycolysis may prevent further differentiation of ESC-derived neurogenic cells to intermediate progenitors. Sox2 was expressed in both naive ESCs and in aPSM cluster 2 (Figure 2I). Sox2 is regulated by pluripotency TFs in ESCs (Young, 2011) and by Pax6 in neurogenic cells (Wen et al., 2008). Thus, we evaluated whether distinct Sox2 regulatory regions may be accessible in naïve ESCs and aPSM cells by scATAC-seq (Figures 2J, 2K, and S2E). Enhancers located ~100 Kb from the Sox2 TSS (SRRs) (Zhou et al., 2014) were accessible in naive ESCs (Figure 2K). Two other regions next to SRRs were also accessible in naive ESCs (Figure 2K, red lines). In aPSM ESCs, all these elements became inaccessible, while increased accessibility was observed at two regions located at ~10 Kb from the Sox2 TSS (Figure 2K). One of these regions, the N1 enhancer, directs Sox2 expression in the posterior neural plate (Takemoto et al., 2006). Another region (R1), located between N1 and SRR enhancers, gained accessibility in aPSM cells (Figure 2K). Thus, besides expressing aPSM genes, ESCs prompted to acquire a PSM-like fate can also activate gene programs observed in endothelial and neurogenic precursors. Neurogenic precursors derived from ESCs recapitulate Sox2 enhancer regulation observed in the embryo's neural plate.

Identification of ESC-derived cell lineages by single-cell omics

ESCs cultured in aPSM-promoting medium were switched to a medium supplemented with hepatocyte growth factor (HGF), insulin growth factor 1 (IGF-1), fibroblast growth factor 2 (FGF-2), LDN193189, and Rspo-3 (HIFLR cells; see STAR Methods) for 48 h and then isolated by FACS based on Pax3-driven GFP expression. Exposure to this medium induces Pax7 activation and, when employed for up to 2 months, can generate multinucleated myofibers, thus recapitulating essential steps of primary myogenesis (Chal et al., 2015). Because of the expected cell heterogeneity induced by the medium employed,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

gene expression and chromatin accessibility were simultaneously evaluated in the same cell by single-cell nuclear RNA-seq (snRNA-seq) and scATAC-seq multiome. Multiome approaches provide a deeper and more accurate characterization of cell types and states compared with integration of scRNA-seq and scATAC-seq datasets obtained in separate experiments (Hao et al., 2021). Individual analysis and visual inspection of snRNA-seq and scATAC-seq datasets indicated high concordance of the two approaches in the identification of specific cell clusters (Figure 3A, snRNA and scATAC panels). Next, we integrated snRNA-seq and scATC-seq measurements by weighted-nearest neighbor (WNN) analysis (Figure 3A, WNN panel). WNN is an unsupervised framework that allows learning the relative utility of simultaneous measurement of multiple modalities (Hao et al., 2021). Clusters obtained by WNN analysis were resolved to generate pseudotime developmental trajectories (Figure 3B). To compute pseudotemporal trajectories, it is necessary to establish the start (root) of the trajectory. Based on WNN analysis, we placed cells with aPSM gene signatures at the root of the trajectory (Figure 3B, root; Table S3). Pseudotemporal ordering identified clusters that were subsequently integrated with WNN (Figure 3C). Binding of TFs to *cis*-regulatory DNA sequences controls gene-expression programs that define cell state and lineages (Davidson, 2006). Therefore, determining simultaneous expression of TFs and enrichment of their cognate DBMs should aid in establishing cell identity. Anterior PSM TFs Meis1 and Pbx1 were broadly expressed and enriched in root clusters (Figure S3B). When unbiasedly queried, the accessible chromatin of root clusters displayed overrepresentation of Meis/Pbx DBMs (Figure S3B). Coexpression of Meis and Pbx and enrichment of their cognate DBMs are consistent with the observation that Meis and Pbx heterodimerize and facilitate chromatin interaction with additional TFs (Knoepfler et al., 1997) (Knoepfler et al., 1999; Berkes et al., 2004; Dell'Orso et al., 2016). Sixteen (16) TFs, expressed in clusters pseudotemporally juxtaposed to the root cluster, are involved in regulating pattern-specific processes (Table S3). Among these TFs, Fli1, which is required for angiogenesis, was also expressed with the endothelial-related genes Flt1 and Flt4 in a small cluster (Figures 3C, endothelial cluster, and S3C). After emerging from root clusters, the trajectory bifurcated to end at two distantly located pseudotemporal clusters (Figures 3B and 3C). Neurogenic basic helix-loop-helix (bHLH) Ascl1 and Nhlh1 were expressed and their DBMs overrepresented and footprinted in neurogenic clusters (Figures 3B–3D). Sox2, NeuroD4, Pax2, Pax8, and Lhx5 were also expressed in this cluster (Figure S3D). The overall gene program of cells hosted in this cluster is indicative of neuronal maturation (neuron projection development, synaptic signaling, and assembly) (Table S3). Lhx5, Pax2, and Pax8 establish a GABAergic inhibitory neurotrasmitter program in the mouse (Pillai et al., 2007).

The other branch of the trajectory ended at a cluster (Figures 3B and 3C, myogenesis cluster composed of 1,127/19,800, 5.6% of total cells analyzed) that hosted cells expressing myogenic bHLH Myod1, Myog, and Myf5 (Figures 3E and S3E; Table S3). In this cluster, there was an overrepresentation of DBMs for Myod1, Myog, and Myf5, and footprinting analysis indicated that these DBMs were occupied (Figures 3E and S3E). Pax7 directs transcription in neuronal precursors and MuSCs (McKinnell et al., 2008; Blake and Ziman, 2014; Lilja et al., 2017; Magli et al., 2019; Zhang et al., 2020). In addition to neurogenic clusters, $Pax7^{+}$ cells coexisted with $Myod1^{+}$ cells (Figure 3F). Cells expressing Myod1

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

or Myogenin were located in the middle section of the myogenic cluster, with Myogenin-positive cells extending further in the cluster (Figure 3G). One tip of the myogenic cluster hosted cells expressing the muscle structural troponin Tnnt2 and Tnnt3 and was devoid of Myod1$^+$ cells (Figures 3F–3H). Thus, the myogenic cluster may recapitulate *in vivo* skeletal myogenesis, where Pax7$^+$/Myod1$^+$ MuSCs cells progressively extinguish Pax7 while maintaining Myod1 expression (activated MuSCs and proliferating myoblasts, Pax7$^-$/Myod1$^+$), initiate to differentiate (myocytes, Pax7$^-$/Myod1$^+$/Myogenin$^+$), and terminally differentiate (Myod1$^-$/Myogenin$^-$/Tnnt3$^+$).

Identification and characterization of genomic regions regulating Pax7 expression in induced ESCs and myogenic lineage

Pax7 regulates neuronal gene expression, including spinal cord neurogenesis, and is critical for maintenance and normal function of MuSCs (Mansouri and Gruss, 1998; Seale et al., 2000; Oustanina et al., 2004; von Maltzahn et al., 2013). As chromatin accessibility can predict gene activation, we inspected the Pax7 locus in ATAC-seq datasets generated at different stages of ESC differentiation. In addition, we wished to compare chromatin accessibility of the Pax7 locus in ESCs and in *bona fide* MuSCs and muscle precursors. For this, we performed ATAC-seq from freshly FACS-isolated MuSCs and from FACS-isolated GFP$^+$ cells from E12.5 embryo-dissected somites obtained by crossing constitutive Pax7$^+$/Cre; Rosa26-YFP mice. Compared with naive and instructed ESCs, aPSM cells displayed increased Pax7 chromatin accessibility, which further augmented in HIFLR cells (Figure 4A). The Pax7 promoter and several regions, including those located at −25 Kb, −3.5 Kb from the TSS, and in the seventh intron (En7), were also accessible in MuSCs and somites (Figure 4A). The transcriptomes of HIFLR cells and E12.5 Pax7-YFP$^+$ somites isolated displayed significant correlation ($R^2 > 0.8$) (Figure S4A). scATAC-seq identified a main cluster (cluster 0) of aPSM cells displaying preferential chromatin accessibility at genes related to anterior/posterior pattern specification and somitogenesis (Figures 4B and S4B, cluster 0). In HIFLR cells, the main cluster (cluster 1) was enriched for accessibility to genes related to muscle structure development (Figures 4B and S4C). Two minor scATAC-seq clusters (clusters 3 and 4) were enriched for accessibility at neurogenic and endothelial genes in both aPSM and HIFLR cells, respectively (Figures 4B, S4B, and S4C). Pax7 was accessible only in the HIFLR myogenic and neurogenic clusters (Figure 4C). Often, when activated, enhancers establish physical contact with the gene promoters that they regulate. We performed *in situ* chromosome conformation capture (Hi-C) (Rao et al., 2014) in aPSM and HIFLR cells and compared them with Hi-C of pluripotent ESCs (Kurup et al., 2020). Hi-C analysis revealed that the Pax7 gene is positioned in a topologically associating domain (TAD) in both ESCs and HIFLR cells (Figure S4D). Despite higher coverage of chromatin interactions, contacts between the Pax7 promoter and En7 element were not detected in pluripotent ESCs. Instead, the En7 region looped to interact with the Pax7 promoter region in both aPSM and HIFLR cells (Figure 4D), indicating that promoter selection by En7 precedes transcriptional activation. Contacts between En7 and the Pax7 promoter were also present in C2C12 cells (Figure 4D). To evaluate their role in a muscle environment, we sought to interfere with endogenous Pax7 −25 Kb, −3.5 Kb, or En7 elements in myogenic C2C12 cells by targeting them with the repressor dCas9-KRAB and specific gRNAs. dCas9-KRAB-mediated perturbation of En7, but not of the −25 or −3.5 Kb element, reduced

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

expression of endogenous Pax7 to levels comparable to those caused by repressing the Pax7 promoter region (Figure 4E). Next, we tested whether En7 behaves as a functional enhancer by cloning it in a luciferase reporter bearing a minimal promoter and stably integrating the resulting construct in the genome of myogenic C2C12 cells. As control, we inserted a genomic region devoid of ATAC-seq peaks, of comparable length, and located –30 Kb from the Pax7 TSS into the luciferase reporter. Compared with either promoter alone or the control reporter, Pax7 En7 elicited robust luciferase activity (Figure 4F), indicating that this genomic region functions as an enhancer. Finally, we interrogated the functional relevance of these DNA elements in directing Pax7 activation by deleting the corresponding accessible (ATAC-seq-defined) regions using a CRISPR-Cas9-based approach. Individual or combined biallelic deletion of the −25 or −3.5 Kb regions in ESCs was without consequence on Pax7 expression (Figures S4E and S4F). In contrast, biallelic deletion of En7 caused a significant reduction (~70%) of Pax7 expression (Figure 4G). These observations were confirmed in two additional independently isolated ESC clones (Figure S4G). Pax7, MyoD, and Myogenin proteins were reduced in En7-deleted cells (Figure 4H). En7 deletion did not interfere with correct Pax7 mRNA splicing (Figure S4I). We further evaluated the role of En7 by conducting RNA-seq in wild-type and En7-deleted cells. The results of these experiments revealed that activation of the myogenic and neurogenic programs was defective in En7-deleted cells (Figures 4I and S4J). Thus, Pax7 En7 controls Pax7 expression and downstream myogenic and neurogenic programs in differentiating ESCs.

DISCUSSION

This study contributes a bulk well as a single-cell resolution resource of both transcriptomes and chromatin accessibility of ESCs induced to differentiate to acquire a PSM-like fate and initial myogenesis and neurogenesis. Exposure to Bmp4 activated gene subsets expressed in ESCs transitioning from naive to primed pluripotency, a state defined as formative pluripotency. Enhancers and promoters employed different syntaxes to activate or repress ESC-induced genes with chromatin accessibility often preceding gene expression or perduring even when transcription subsided. Single-cell approaches permitted the identification of heterogeneous cell types including myogenic and neurogenic precursors. Gene signatures detected in instructed ESCs suggest that neuromesodermal progenitors may participate to the specification of both neurogenic and myogenic precursors (Henrique et al., 2015; Gouti et al., 2017). Neural progenitors and differentiated neurons were observed during the initial phases of directed myogenic differentiation of human pluripotent stem cells (Xi et al., 2020). Underlying species difference, neurogenic and myogenic cells emerged 3 and 4 weeks after directed differentiation of human pluripotent stem cells (hiPSCs), respectively (Xi et al., 2020) whereas the data reported here indicate that mouse pluripotent stem cells give rise to neurogenic and myogenic cells 8 days after initial ESCs induction. In two other studies, hiPSCs activated the myogenic program after 10–15 days, though with different PAX7 activation kinetics (Wu et al., 2018; Al Tanoury et al., 2020). The differences in neurogenic and myogenic induction and myogenic activation kinetics are likely due to different protocols employed in different studies. Given the relevant role played by Pax7 in neurogenesis and myogenesis, we have leveraged our datasets to identify Pax7 upstream regulatory regions directing expression in both cell lineages. Physical interaction of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Pax7 upstream regulatory regions with the cognate promoter preceded Pax7 expression in differentiating ESCs and was observed to occur also in C2C12 myogenic cells. The enhancer region we have identified is distinct from those directing Pax7 expression in mouse cranial neural crest, facial mesenchyme, mesencephalon, pontine reticular nucleus (Lang et al., 2003), and from the enhancer that directs Pax7 expression in chick embryo neural crest (Vadasz et al., 2013). We anticipate that integration of the datasets reported here to characterize Pax7 regulatory regions may be employed to query the dynamics of gene-regulatory networks occurring in ESCs induced to acquire a PSM-like fate and initial myogenic or neurogenic differentiation.

Limitations of the study

Mouse ESCs represent a culture model and do not reflect the complex series of events occurring during development. Transcriptomes and epigenomes of ESCs only partially recapitulate those observed in the embryo. Moreover, our study is limited to four ESC time points (naive ESCs, Bmp4-instructed ESCs, aPSM, and HIFLR cells). A more detailed time course evaluation of ESCs induced to acquire a PSM fate and longer exposure of PSM cells to differentiating factors would add further granularity to our datasets. An additional limitation of this study relates to the composition, doses, and timing of cells' exposure to defined factors and small molecules, which determine the outcome of the observations.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for reagents should be directed to Lead contact Vittorio Sartorelli (vittorio.sartorelli@nih.gov).

Materials availability—The plasmids and cells generated in this study are available from the lead contact, upon request.

Data and code availability

- RNA-seq, ATAC-seq, H327ac and H3K4me1 ChIP-seq, HiC, Single cell RNA-seq, Single cell ATAC-seq and Multiome single nuclei ATAC and gene expression data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Original qPCR data and Western blot images will be shared by the lead contact upon request.
- All original data have been deposited on GEO (Gene Express Omnibus) repository and are publicly available as of the date of publication.
- Original codes have been deposited at GitHub and are publicly available as of the date of publication. DOIs are listed in key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines—All cells were cultured at 37°C with 5% CO_2. HEK293T and C2C12 cells (ATCC) were grown in 1× DMEM supplemented with 10% and 20% of qualified fetal bovine serum (FBS) (GIBCO), respectively. Mouse Embryonic Stem Cells (ESCs) expressing Pax3-GFP are described in (Chal et al., 2015). ESCs were cultured as follows:

Maintenance.: ESCs were plated on 0.1% gelatin (Millipore) and cultured in DMEM (Gibco) supplemented with 15% inactivated fetal bovine serum (Hyclone), 1% penicillin-streptomycin, 2 mM L-glutamine, 0.1 mM non-essential amino acids, 0.1% β-mercaptoethanol, 1,500 U/ml LIF and 2i inhibitors (Stemgent).

Serum-free differentiation of mouse ESCs.: ESCs differentiation was induced as described (Chal et al., 2015) with minor modifications. Briefly, ESCs were trypsinized and plated on gelatin-coated plates in serum-free N2B27 medium (N2B27,1% Knock-out Serum Replacement (KSR, Gibco), 0.1% bovine serum albumin (Gibco) and BMP4 (Peprotech) at 10 ng/ml. After 2 days, cells were shifted to RDL medium, (DMEM, 1%FBS, 14% KSR,10 ng/ml Rspo3 (Peprotech, R&D Biosystems), 0.5% DMSO (Sigma), and 0.1 μM LDN193189 (Stemgent), and cultured for 4 additional days. Subsequently, medium was changed to DMEM, 1%FBS, 14% KSR, 0.1% BSA supplemented with 10 ng/ml Rspo3, 10 ng/ml HGF, 2 ng/ml IGF-1, 20 ng/ml FGF-2 (Peprotech, R&D Biosystems) and 0.1 μM LDN193189 (HIFLR medium).

Mice and animal care—Mice were housed in a pathogen-free facility, all the experiments were performed according to the National Institutes of Health (NIH) Animal Care and Use regulations. $Pax7^+$/Cre; Rosa26-YFP embryos (E12.5) were generated by timed mating between $Pax7^+$/Cre (Keller et al., 2004) and Rosa26-YFP (Rosa26-YFPY/Y) mice (Srinivas et al., 2001). All animal protocols have been revised and approved by the NIAMS Animal Care and Use Committee (ACUC).

Bacterial strains—DH5α and TOP10 cells were obtained from New England BioLabs (NEB) and Thermo Fisher Scientific Inc. Both DH5α and TOP10 bacterial strains were grown in LB medium at 37°C and used to propagate plasmids.

METHODS DETAILS

Single guide RNA (sgRNA) design and Cas9 vector assembly—sgRNA for CRISPR/Cas9 assay were designed using *CHOPCHOP* web tool (https://chopchop.cbu.uib.no/). sgRNAs were tested by T7 Endonuclease assay (https://www.neb.com/products/m0302-t7-endonuclease) to evaluate editing efficiency, and selected sgRNAs were cloned into the Cas9 vector pSpCas9 (BB)-2A-GFP (pX458; Addgene plasmid#48138) or into the dCAS9-KRAB repressor GFP-plasmid (addgene#71236) (Primers are reported in Table S4).

Plasmids Construction—The genomic regions corresponding to the En7 enhancer (chr4:139771035–139771985) and a control region (chr4:139,863,500–139,864,500.) were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

amplified from genomic DNA and cloned into the 5' XhoI and 3' EcorV sites in pGL4.26[luc2/minP/Hygro] vector (Promega, #E8441). Primers are reported in Table S4).

Transfections and genome editing—Plasmid transient transfections were performed using Lipofectamine 2000 (Invitrogen). For genome editing, ESCs cells were transfected in suspension with 1.5μg of pSpCAS9-GFP with different sgRNAs and plated on 6 well gelatin-coated plates. 48 h after transfection, cells were seeded in 100mm culture dishes at a density of 1.5 and 10 cells/ml; 2 weeks later, clones were collected and screened for genomic editing. For luciferase reporter assay, 2×10^5 C2C12 cells were transfected with ~1–3 μg of Pax7 luciferase reporter vectors and plated in 6 well plates. 72 h post-transfection medium was changed and cells were selected with 300μg/ml Hygromycin (Sigma, 31282–04-9) for 5 days.

Lentiviral production—Lentiviral particles were generated using HEK293T cells (ATCC). Cells were seeded at 50–60% confluency in 100 mm cell culture dish. The following day, cells were transfected with 5μg of the vector of interest (dCAS9-KRAB with specific sgRNAs), 1μg of pMD2.G and 4μg of psPAX2 (Addgene) using 30μL of Lipofectamine (2000). 24 h post-transfection the media was changed. The supernatant was collected 48hr post-transfection, filtered with a 0.45-mm PVDF filter (Millipore), and concentrated using the lenti-x concentrator (Takara, 631232) according to the manufacturer's instructions. Concentrated viruses were stored at −80°C.

C2C12 cells were transduced in suspension and plated in DMEM medium supplemented with 6 μg polybrene. Cells were collected five days post-transduction and GFP positive cells were sorted for RNA extraction.

Luciferase reporter assay—Luciferase activity was detected by adding luciferin (Promega, E1500) to cell extracts of transfected C2C12 cells as a substrate at a final concentration of 0.15 mg/mL according to the manufacturer's instructions. Fluorescence was detected and quantified on a luminometer (Synergy™ H1 microplate reader, BioTek). Luciferase activity was normalized by total protein concentration.

ESCs FACS sorting—ESCs were trypsinized and FACS-isolated gating on positive GFP and DAPI staining, and negative PI staining using FACSAria fusion (BD Biosciences)

Somites isolation and FACS sorting—Somites were isolated from E12.5 $Pax7^+$/Cre; Rosa26-YFP embryos. Cells were isolated from regions corresponding to posterior to forelimb area or mid-trunk area. 20 to 25 pairs of somites were obtained from each embryo. The spinal cord and lateral paraxial mesoderm were separated by sharp forceps to avoid sorting of YFP^+ dorsal spinal cord cells. The separated lateral tissue was immediately placed into 500ul of 2.5mg/ml Papain/1xPBS (Sigma-Aldrich), minced with scissors and incubated at 33°C for 20 min for disassociation. 500ul of sorting medium (15% FBS in 1x PBS) was added to the digested tissue to stop the reaction. The obtained tissue was diluted in 10mL of sorting medium (15% FBS in 1x PBS), then filtered with a 0.45mm filter (Millipore). Cells were centrifuged at 1300 rpkm for 5 min and resuspended in 500ul of Sorting medium. Final

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

FACS-sorting was conducted on a BD FACSAria IIIu machine by gating the YFP channel. About 100,000–200,000 YFP^+ somitic cells were collected from one sorted embryo.

Muscle dissection and MuSC FACS-Isolation—Hindlimb muscles from 3-month-old C56BL/6J wild-type mice were dissociated into single cells by enzymatic digestion and live cells were isolated by FACS. Skeletal MuSCs were sorted following described methods (Liu et al., 2015). Briefly, hindlimb muscles from 3-month-old adult wild-type mice were minced and digested with collagenase for 1 h and MuSCs were released from muscle fibers by further digesting the muscle slurry with collagenase/dispase for an additional 30 min. After filtering out the debris, cells were incubated with the following primary antibodies: biotin anti-mouse CD106 (anti-VCAM1, BioLegend 105704; 1:75), PE/Cy7 Streptavidin (BioLegend 405206; 1:75), Pacific Blue anti-mouse Ly-6A/E (anti-Sca1, BioLegend 108120; 1:75), APC anti-mouse CD31 (BioLegend 102510; 1:75) and APC anti-mouse CD45 (BioLegend 103112; 1:75). Satellite cells were sorted by gating VCAM1-positive, Pacific Blue-labeled Sca1-negative, and APC-labeled CD31/CD45-negative cells. SYTOX Green (ThermoFisher Scientific S7020; 1:30,000) was used as a counterstain.

Antibodies—Western blot experiments were conducted using the following antibodies: anti-Pax7 (DSHB), anti-Myod1 (Santa Cruz. 71629), anti-MyoG (Santa Cruz. 12732), anti-beta-Actin (Santa Cruz. 130656), anti-Otx2 (cell signaling. mAb #11943), and anti-Nanog (Santa Cruz, sc-374001). Goat anti-Rabbit IgG-HRP or goat anti-mouse IgG-HRP (Azure biosystem, AC2114 and 2115 respectively) were used as secondary antibodies for immunoblotting. Antibodies anti-H3K27Ac (ab4729, Abcam) and anti-H3K4me1(ab8895, Abcam) were employed in ChIP-seq experiments.

Quantitive and Semi-Quantitive PCR—Total RNA was extracted using TRIzol Reagent (Invitrogen) according to the manufacturer's protocol. cDNAs were synthesized with the qScript cDNA kit (Quanta) containing random primers. Reverse transcribed cDNA was diluted 10 times and SYBR green real-time PCR was performed on the Applied Biosystems StepOne Plus Real-Time PCR system. Quantifed mRNA levels were normalized to 18S and relative expressions were calculated according to ΔΔCt. A full list of primers is reported in Table S4.

To screen for positive CRISPR/CAS9 clones carrying the expected deletions, genomic DNA was extracted using a Blood & Tissue Extraction kit (Qiagen), and PCR was performed on Applied Biosystems PCR system. Primers are reported in Table S4.

cDNAs were retrotranscribed from RNA of either WT or En7-deleted ESC clones and Pax7 exons 7 and 8 were Sanger-sequenced.

Immunoblotting—Cells were lysed in lysis buffer [100 mM Tris-HCl pH 7.5 and 5% sodium dodecyl sulfate (SDS)]. Proteins (20–30 μg) were incubated at 95°C for 5 min, resolved by SDS-PAGE, and transferred to nitrocellulose membranes. Subsequently, membranes were incubated with blocking solution (0.1% Tween 20, 5% non-fat milk in 1XPBS) for 1 h at room temperature, and primary antibodies were added overnight at 4°C. Membranes were then washed with 1XPBS/0.05% Tween 20 and incubated with secondary

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

antibody (Azura HRP Ab, CatN X) for 1 h at room temperature. Bioanalytical Imaging System-600 (Azura Biosystem, Inc) was used for protein visualization.

ChIP-seq—4×10^6 ESCs were used for chromatin immunoprecipitation. Cells were crosslinked in 1% formaldehyde and processed according to published protocols (Mousavi et al., 2013). Briefly, cells were lysed in RIPA buffer (1× PBS, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) and centrifuged at 2,000 rpm for 5 min. The chromatin fraction was shared by sonication (four times, each lasting 30). The resulting sheared chromatin samples were immunoprecipitated overnight, and washed in buffer I (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), buffer II (20 mM Tris-HCl [pH 8.0], 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), buffer III (10 mM Tris-HCl [pH 8.0] 250 mM LiCl, 1% NP-40; 1% sodium deoxycholate, 1 mM EDTA), and Tris-EDTA (pH 8.0). All washes were performed at 4°C for 5 min. Finally, crosslinking was reversed in elution buffer (100 mM NaHCO3, 1% SDS) at 65°C overnight. For ChIP-seq, 10 ng immuno-precipitated DNA fragments were used to prepare ChIP-seq libraries with the NEBNext Ultra II DNA library prep kit for Illumina (New England Biolabs, (#E7645) following the manufacturer's protocol. The libraries were sequenced on a NextSeq550 or NovaSeq6000 Illumina instrument.

ATAC-seq—ATAC-seq was performed according to a published protocol (Buenrostro et al., 2013) with minor modifications. Briefly, 5×10^4 cells were washed with 50ul of 1xPBS and lysed in 50ul of Lysis Buffer (10mM Tris-HCl, pH7.4, 10mM NaCl, 3mM MgCl2, 0.1% of IGEPAL CA-630). To tag and fragment accessible chromatin, nuclei were centrifuged at 500x g for 10min and resuspended in 40ul of transposition reaction mix with 2ul Tn5 transposase (Illumina Cat# FC-121–1030). The reaction was incubated at 37°C with shaking at 300rpm for 30min. DNA fragments were then purified and amplified by PCR (12–15 cycles based on the amplification curve). Purified libraries were sequenced on NextSeq550 or NovaSeq6000 Illumina instrument.

RNA-seq—For transcriptome analysis (RNA-seq), poly $(A)^+$ mRNA libraries were generated in triplicate using NEBNext Ultra II RNA library preparation kit for Illumina (NEB #E7490) according to the manufactory instructions.

Single Cell RNA-Seq—ESCs were collected, washed once with 2×PBS, and re-suspended in PBS with 0.04% bovine serum albumin. Cellular suspensions were loaded on a Chromium Instrument (10x Genomics) to generate single-cell GEMs. Single-cell RNA-seq libraries were prepared using a Chromium Single Cell 3' Library & Gel Bead Kit v3.1 (P/N 1000121, 10x Genomics). GEM-RT was performed in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad; P/N 1851197): 53°C for 45 min, 85°C for 5 min; held at 4°C. Following retrotranscription, GEMs were broken and the single-strand cDNA was purified with DynaBeads MyOne Silane Beads. cDNA was amplified using the C1000 Touch Thermal cycler with 96-Deep Well Reaction Module: 98°C for 3 min; cycled 12 times: 98°C for 15 s, 63°C for 20 s, and 72°C for 1 min; 72°C for 1 min; held at 4°C. Amplified cDNA product was purified with the SPRIselect Reagent Kit (0.6× SPRI). Indexed sequencing libraries were constructed using the reagents in the Chromium

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Single Cell 3' Library & Gel Bead Kit v2, following these steps: (1) end repair and A-tailing; (2) adaptor ligation; (3) post-fragmentation, end repair, and A-tailing double size selection cleanup with SPRI-select beads; (4) sample index PCR and cleanup. The barcode sequencing libraries were diluted at 3 nM and sequenced on an Illumina NovaSeq6000 using the following read length: 28bp for Read1, 8 bp for I7 Index, and 91 bp for Read2.

Single Cell ATAC-Seq—Cells were washed twice with 1xPBS with 0.04%BSA (Sigma) and single-cell ATAC was performed using Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit v1.1 (10x genomics, 1000175). Briefly, 1X106 cells were lysed using chilled lysis buffer (10mM Tris HCl, 10mM NaCl, 3mM MgCl2, 0.1% Tween 20, 0.1% Nonidet P40 substitute, 0.01% Digitonin and 10% BSA). Following lysis, nuclei were washed with wash buffer (10mM Tris HCl, 10mM NaCl, 3mM MgCl2, 0.1% Tween 20 and 1% BSA) and re-suspended in Nuclei buffer (10x Genomics, PN-2000153). Subsequently, nuclei suspensions were incubated with transposition mix for 1 h at 37C in a C1000 Touch Thermal cycler. Nuclei were diluted in Diluted Nuclei Buffer according to 10X Genomics recommendations (10X Genomics) and loaded on a Chromium Instrument (10x Genomics) to generate GEMs. Samples were incubated in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad; P/N 1851197) programmed as 72°C for 5 min; 98°C for 30 s, cycled 12 times: 98°C for 10 s, and 59°C for 30s; 72°C for 1 min; held at 4°C. Amplified products were purified with Dynabeads MyOne Silane beads followed by SPRIselect Clean-up. Indexed sequencing libraries were constructed using the reagents in the Chromium Single Cell ATAC Reagent Kit v1.1 to add the P7 and P5 sequences used in Illumina bridge amplification, and a sample index. Final libraries were diluted to 3 nM and sequenced on an Illumina NovaSeq6000 using the following read length: 50 bp for Read1, 8 bp for I7 Index, 15 bp for i5 Index, and 50 bp for Read2.

Chromosome conformation capture-Hi-C—Hi-C experiments were performed using the Arima-HiC kit (A510008GFP). Briefly, 2×10^6 cells were crosslinked with 2% formaldehyde (37%, Sigma) for 10 min followed by incubation with 200mM glycine (sigma) to stop cross-linking (sigma). Cells were washed three times with 1xPBS and resuspended in 20ul elution buffer (Arima). Cells were then lysed and chromatin was extracted, digested, and biotinylated following Arima-HiC kit instructions. Biotin-labeled DNA was quantified and fragmented to 400 bp length using Covaris (ME220) About 300ng of DNA was used for biotin enrichment. Final DNA libraries were constructed using the Swift Biosciences kit for Library Preparation (Accel-NGSO 2S Plus DNA Doc A160140 v00).

Single Cell Multiome—Single-cell ATACseq and gene expression omics were performed using Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (10x Genomics kit, CG000338) according to the manufacturer's instruction. Briefly, cells were washed with 1xPBS and resuspended in PBS-0.04%BSA. Nuclei were isolated by incubating the cells in chilled Lysis Buffer (10mM Tris HCl, 10mM NaCl, 3mM MgCl2, 0.1% Tween 20, 0.1% Nonidet P40 substitute, 0.01% Digitonin and 10% BSA) for 3 min. After washing, nuclei suspensions were incubated in a Transposition Mix that includes a Transposase and loaded into the Chromium instrument (10X Genomics) for GEMs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

generation. Samples were then incubated in a C1000 Touch Thermal cycler with 96-Deep Well Reaction Module (Bio-Rad; P/N 1851197) programmed as 37°C for 45 min; 25°C for 30 min, 4°C holds. Amplified products were purified with Dynabeads MyOne Silane beads followed by SPRIselect clean up. Eluted samples containing barcoded transposed DNA and barcoded cDNA were pre-amplified via PCR and cleaned up using SPRIselect. Eluted pre-amplified reactions were used to generate ATACseq and gene expression (GEX) libraries according to the 10X Genomics protocol. 1ml of each library was run on High sensitivity D1000 TapeStation to determine the correct average fragment size. Libraries were diluted at 3 nM and sequenced on an Illumina NovaSeq6000 using the following read length: 50bp for Read1, 8 bp for I7 Index, 24bp for i5 index, and 49bp for Read2.

Gene Ontology analysis—Gene Ontology (GO) analysis of genes from different experiments was performed with Metascape, gene annotation, and analysis resource. (https://metascape.org/gp/index.html#/main/step1).

Bulk RNA-Seq analysis—RNA-Seq data were generated using Illumina NovaSeq6000 system. Raw sequencing data were processed with bcl2fastq/2.20.0 to generate FastQ files. Adapter sequences were removed using trimgalore/0.6.6 (https://github.com/FelixKrueger/TrimGalore). Single-end reads of 50 bases were mapped to the mouse transcriptome and genome mm10 using TopHat 2.1.1 (Trapnell et al., 2009). Gene expression values (RPKM: Reads Per Kilobase exon per Million mapped reads) were calculated using Partek Genomics Suite 7.18, which was also used for the PCA and ANOVA analyses.

ChIP-seq and ATAC-Seq analysis—Sequencing data were generated with an Illumina NovaSeq6000 system. FastQ files were generated with bcl2fastq/2.20.0. Adapter sequences were removed using trimgalore/0.6.6. Reads of 50 bases were aligned to the mouse genome build mm10 with Bowtie/1.1.1 (Langmead et al., 2009), allowing two mismatches. Uniquely mapped and non-redundant reads were used for peak calling using MACS 1.4.2 (Zhang et al., 2008) with a p value cutoff of 1.0E-05. Histone model was applied for CHIP-seq samples while transcription factor model for ATAC-seq samples. Only regions called in both replicates were used in downstream analysis. Bigwig files were generated with BedGraphToBigWig (Kent et al., 2010) and Bedtools/2.29.2 (Quinlan and Hall, 2010). Peak intensities were normalized as tags per 10 million reads (RP10M) in the original library. Peaks were assigned to the closest TSS with HOMER.

Single Cell RNA-Seq analysis—Demultiplexing and reads alignment were performed with CellRanger ver. 3.1.0 (10x Genomics) with default parameters. For all analyses, we employed standard pre-processing for all single-cell RNA-seq datasets. Filtering, variable gene selection, and dimensionality reduction were performed using the Seurat ver.4.1.0 (Butler et al., 2018).

Filtering

1. Cells were filtered out based on the following criteria:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- To eliminate low quality cells and debris, cells with less than 200 detected genes and cells with more than 20% of UMIs mapped to mitochondrial genes were excluded.
- To remove potential cell doublets and aggregates only cells showing a nFeature_RNA between 1600 and 8000 were selected.

Normalization

1. For each cell, UMI counts per million were log-normalized using the natural logarithm.
2. In each dataset, we aimed to identify a subset of features (e.g., genes) exhibiting high variability across cells to prioritize downstream analysis. Variable genes were selected applying thresholds calculated using binned values from log average expression and dispersion for each gene.

Clustering

1. The expression level of highly variable genes was scaled along each gene and cell-cell variation was regressed out by number of detected molecules, and mitochondrial gene expression.
2. Data were projected onto a low-dimensional subspace of PCA (principal component analysis) using dimensional reduction. The number of PCA was decided through assessment of statistical plots.
3. Cells were clustered using a graph-based clustering approach optimized by the Louvain algorithm with resolution parameters and visualized using two-dimensional UMAP (Uniform Manifold Approximation and Projection).
4. To define cluster identity, expression and distribution of known markers were evaluated.

Integration of scRNA-Seq datasets—We integrated scRNA-seq datasets of naive and instructed ESCs using Harmony (v1.0) (Korsunsky et al., 2019) for batch correction. Harmony was applied on the first 30 PCA components (RunHarmony function setting assay.use to RNA, max.iter.harmony to 10, max.iter.cluster to 20, sigma to 0.1). After dimensional reduction, cells were clustered by the Louvain algorithm (resolution at 0.25,1.0 and 2.0), and gene expression distributions were visualized using two-dimensional UMAP Harmony.

Pseudotemporal Analyses—Pseudotime was calculated by the Monocle3 R package (Cao et al., 2019) following the conversion of gene expression data and cell metadata including cell type labels from the Seurat object to Monocle object. After passing parameters in Seurat to the “residual_model_formula_str” argument in the “preprocess_cds” function, significant PCs were selected to further reduce the data dimensionality using uniform manifold approximation (UMAP). The beginning of pseudotime was selected in the UMAP dimension and the distribution of the pseudotime was calculated by “order_cells”

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

function. Genes differentially expressed across a single cell trajectory were identified with "graph_test" and visualized by "plot_cells" function.

Single Cell ATAC-Seq analysis—Demultiplexing and reads alignment were performed using Cellranger-atac ver. 1.1.0 with default parameters). For all analyses, we employed standard pre-processing for all single-cell ATAC-seq datasets. Filtering, variable gene selection, and dimensionality reduction were performed using the Signac ver.1.5.0 (Stuart et al., 2019).

Filtering: ESCs with fewer than 3 detected genes were excluded. Additional standard QC steps (duplicates removal, evaluation of number of fragments per cell and number of fragments per peak, fraction of reads mapping to blacklist regions, nucleosome pattern signal, and transcriptional start site (TSS) enrichment) were also applied.

Clustering

1. LSI dimensionality reduction was applied to all samples (RunTFIDF function with method equal to 2, FindTopFeatures function setting min.cutoff to q0, and RunSVD function using the peaks as assay) and data were projected onto a low-dimensional sub-space of PCA (principal component analysis) using dimensional reduction.
2. Cells were clustered using a graph-based clustering approach optimized by the Louvain algorithm with resolution parameters and visualized using two-dimensional UMAP (Uniform Manifold Approximation and Projection).
3. Gene expression in each cluster was predicted employing GeneActivity function, obtained gene activity was generated in the Seurat object (normalization method with log-normalized and scale factor with the median of the sample).

Downstream analysis of scATAC-Seq data—Downstream analysis was performed using R 4.1.0 applying Signac ver.1.5.0 (Stuart et al., 2021). After correction of GC bias using RegionStats function in Sinac, peaks correlated with the expression of nearby genes were identified with LinkPeaks function. Next, the tracks of ATAC-seq peaks for each cluster were visualized using CoveragePlot function (−0.5Kb/+14Kb from the TSS). Transcriptional factor (TF) activities on the ATAC-seq data were calculated using the Signac implementation of TFBSTools (Tan and Lenhard, 2016) with JASPAR2020 vertebrates TF binding models database (746 TFs) (Fornes, Oriol et al., 2019). The footprinting of each TF near the TSS regions was visualized using Footprint and PlotFootprint functions.

Integration of scATAC-Seq data—scATAC-seq datasets were integrated as described (Stuart et al., 2019). Common anchors in the scATAC-seqs datasets were identified by Seurat function FindTransferAnchors with Latent Semantic Indexing (LSI) reduction and the first 40 components were calculated. These components were then used to generate a Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction. Nearest-Neighbor graph was generated after excluding the first component following

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the standard guidelines from the Satija lab (https://satijalab.org/signac/). Cells were then clustered with the Seurat's Louvain algorithm.

Integration of scRNA-seq and scATAC-seq data—scATAC-seq and scRNA-seq data were integrated as described (Stuart et al., 2019). Using scRNA-seq and scATAC-seq data, cluster labels in the scATAC-seq dataset were predicted by the Seurat function FindTransferAnchors on the Canonical Correlation Analysis (CCA) space and were trasnferred to the scATAC-seq dataset. This operation used the variable features of the scRNA-seq analysis on the RNA assay as the reference data and the gene activity matrix of the scATAC-seq analysis as the query data. After transferring the labels through TransferData function, they were merged into the two Seurat objects, and co-visualized with clusters labeled by the scRNAseq cluster labels. Finally, dot-plots visualizing averaged gene expressions and chromatin accessibility were generated by calculating z-scores.

Single Cell Multiome analysis—The multi-omic dataset was realigned to mm10 using cellranger-arc version 2.0.0 (10x Genomics). The resulting RNA count matrix was filtered for cells with between 1,200 and 10,000 reads while ATAC peak matrix was filtered with between 100 and 10,000 reads. Next, the WNN graph was constructed to learn the relative utility of each data modality in each cell (https://satijalab.org/seurat/articles/weighted_nearest_neighbor_analysis.html:

WNN analysis of 10x Multiome, RNA + ATAC). The R packages Seurat ver.4.1.0 and Signac ver.1.5.0 were used for data scaling, transformation, clustering, dimensionality reduction, differential expression analysis, and most visualizations. Pseudotime calculation and visualization are described in detail in the "Pseudotemporal Analyses" section.

Hi-C data analysis—Hi-C datasets were processed using Juicer 1.6 (Durand et al., 2016) Raw reads were aligned to the mm10 reference genome using BWA (Li and Durbin, 2010). Reads that aligned to more than two places in the genome were discarded. The remaining aligned reads were filtered based on mapping quality score (MAPQ <30). Contact matrices were generated at different base pair resolutions ranging from 1MB to 5kb. Downstream analysis was performed with HOMER (Heinz et al., 2010). The Hi-C summary output from Juicer was used to generate a paired end tags using the HOMER makeTagDirectory script. Extraction of significant interactions was done using HOMER at 10kb resolution with a 25kb window, by running the HOMER AnalyzeHiC script. ATAC-Seq peak regions for each sample were used to assess overlapping of significant interactions with regions of open chromatin. HiCExplorer 3.6 (Wolff et al., 2020) was employed to visualize significant interactions along with ATAC-seq tracks for each sample at defined regions of interest. The code executed in the described pipeline is available in Key Resources Table of STAR Methods.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data from quantitative PCR are expressed as mean ± standard error from three different experiments (n = 3). Significant differences were analyzed by the two-tailed, unpaired, Student's t-test and values were considered significant at $p < 0.05$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank the NIAMS Genomic Technology, Biodata Mining and Discovery, Flow Cytometry, and Light Imaging Sections. Dr. Hong-Wei Sun (Biodata Mining and Discovery Section) provided useful suggestions for data analysis. This study utilized the high-performance computational capabilities of the Helix Systems at the NIH, Bethesda, MD, USA (https://helix.nih.gov/). This work was supported in part by the Intramural Research Program of the NIAMS at the NIH (grants AR041126and AR041164 to V.S.).

REFERENCES

Acampora D, Omodei D, Petrosino G, Garofalo A, Savarese M, Nigro V, Di Giovannantonio LG, Mercadante V, and Simeone A. (2016). Loss of the otx2-binding site in the Nanog promoter affects the integrity of embryonic stem cell subtypes and specification of inner cell mass-derived epiblast. Cell Rep. 15, 2651–2664. 10.1016/j.celrep.2016.05.041. [PubMed: 27292645]

Al Tanoury Z, Rao J, Tassy O, Gobert B, Gapon S, Garnier JM, Wagner E, Hick A, Hall A, Gussoni E, and Pourquié O. (2020). Differentiation of the human PAX7-positive myogenic precursors/satellite cell lineage in vitro. Development 147, dev187344. 10.1242/dev.187344.

Allis CD, and Jenuwein T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. 17, 487–500. 10.1038/nrg.2016.59. [PubMed: 27346641]

Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, and Chambers I. (2018). Functional dissection of the enhancer repertoire in human embryonic stem cells. Cell Stem Cell 23, 276–288.e8. 10.1016/j.stem.2018.06.014. [PubMed: 30033119]

Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, and Tapscott SJ (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol. Cell 14, 465–477. [PubMed: 15149596]

Blake JA, and Ziman MR (2014). Pax genes: regulators of lineage specification and progenitor cell maintenance. Development 141, 737–751. 10.1242/dev.091785. [PubMed: 24496612]

Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, Nichols J, Smith A, and Bertone P. (2015). Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. Dev. Cell 35, 366–382. 10.1016/j.devcel.2015.10.011. [PubMed: 26555056]

Brack AS, and Rando TA (2012). Tissue-specific stem cells: lessons from the skeletal muscle satellite cell. Cell Stem Cell 10, 504–514. 10.1016/j.stem.2012.04.001. [PubMed: 22560074]

Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T, and Wysocka J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell 14, 838–853. 10.1016/j.stem.2014.04.003. [PubMed: 24905168]

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, and Greenleaf WJ (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218. 10.1038/nmeth.2688. [PubMed: 24097267]

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. 10.1038/nbt.4096. [PubMed: 29608179]

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502. 10.1038/s41586-019-0969-x. [PubMed: 30787437]

Catarino RR, and Stark A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. 32, 202–223. 10.1101/gad.310367.117. [PubMed: 29491135]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Chal J, Al Tanoury Z, Oginuma M, Moncuquet P, Gobert B, Miyanari A, Tassy O, Guevara G, Hubaud A, Bera A, et al. (2018). Recapitulating early development of mouse musculoskeletal precursors of the paraxial mesoderm in vitro. Development 145. 10.1242/dev.157339.

Chal J, Oginuma M, Al Tanoury Z, Gobert B, Sumara O, Hick A, Bousson F, Zidouni Y, Mursch C, Moncuquet P, et al. (2015). Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. Nat. Biotechnol. 33, 962–969. 10.1038/nbt.3297. [PubMed: 26237517]

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, and Zaret KS (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell 9, 279–289. 10.1016/s1097-2765(02)00459-8. [PubMed: 11864602]

Coffinier C, Thépot D, Babinet C, Yaniv M, and Barra J. (1999). Essential role for the homeoprotein vHNF1/HNF1beta in visceral endoderm differentiation. Development 126, 4785–4794. 10.1242/dev.126.21.4785. [PubMed: 10518495]

Cornacchia D, Zhang C, Zimmer B, Chung SY, Fan Y, Soliman MA, Tchieu J, Chambers SM, Shah H, Paull D, et al. (2019). Lipid deprivation induces a stable, naive-to-primed intermediate state of pluripotency in human PSCs. Cell Stem Cell 25, 120–136.e10. 10.1016/j.stem.2019.05.001. [PubMed: 31155483]

Costello I, Nowotschin S, Sun X, Mould AW, Hadjantonakis AK, Bikoff EK, and Robertson EJ (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. Genes Dev. 29, 2108–2122. 10.1101/gad.268979.115. [PubMed: 26494787]

Dale L, Howes G, Price BM, and Smith JC (1992). Bone morphogenetic protein 4: a ventralizing factor in early Xenopus development. Development 115, 573–585. 10.1242/dev.115.2.573. [PubMed: 1425340]

Darabi R, Arpke RW, Irion S, Dimos JT, Grskovic M, Kyba M, and Perlingeiro RCR (2012). Human ES- and iPS-derived myogenic progenitors restore DYSTROPHIN and improve contractility upon transplantation in dystrophic mice. Cell Stem Cell 10, 610–619. 10.1016/j.stem.2012.02.015. [PubMed: 22560081]

Davidson EH (2006). The Regulatory Genome: Gene Regulatory Networks in Development and Evolution (Academic Press).

Dell'Orso S, Wang AH, Shih HY, Saso K, Berghella L, Gutierrez-Cruz G, Ladurner AG, O'Shea JJ, Sartorelli V, and Zare H. (2016). The histone variant MacroH2A1.2 is necessary for the activation of muscle enhancers and recruitment of the transcription factor Pbx1. Cell Rep. 14, 1156–1168. 10.1016/j.celrep.2015.12.103. [PubMed: 26832413]

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, and Aiden EL (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98. 10.1016/j.cels.2016.07.002. [PubMed: 27467249]

Evano B, and Tajbakhsh S. (2018). Skeletal muscle stem cells in comfort and stress. NPJ Regen. Med. 3, 24. 10.1038/s41536-018-0062-3. [PubMed: 30588332]

Field A, and Adelman K. (2020). Evaluating enhancer function and transcription. Annu. Rev. Biochem. 89, 213–234. 10.1146/annurev-biochem-011420-095916. [PubMed: 32197056]

Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranasic D, et al. (2020). JASPAR 2020: Update of the open-access database of transcription factor binding profiles. Nucleic. Acids Res. 48, D87–D92. 10.1093/nar/gkz1001. [PubMed: 31701148]

Fuchs E, and Blau HM (2020). Tissue stem cells: architects of their niches. Cell Stem Cell 27, 532–556. 10.1016/j.stem.2020.09.011. [PubMed: 33007238]

Gómez-López S, Wiskow O, Favaro R, Nicolis SK, Price DJ, Pollard SM, and Smith A. (2011). Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells in vitro. Glia 59, 1588–1599. 10.1002/glia.21201. [PubMed: 21766338]

Gong J, Gu H, Zhao L, Wang L, Liu P, Wang F, Xu H, and Zhao T. (2018). Phosphorylation of ULK1 by AMPK is essential for mouse embryonic stem cell self-renewal and pluripotency. Cell Death Dis. 9, 38. 10.1038/s41419-017-0054-z. [PubMed: 29348566]

Götz M, Stoykova A, and Gruss P. (1998). Pax6 controls radial glia differentiation in the cerebral cortex. Neuron 21, 1031–1044. 10.1016/s0896-6273(00)80621-2. [PubMed: 9856459]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Gouti M, Delile J, Stamataki D, Wymeersch FJ, Huang Y, Kleinjung J, Wilson V, and Briscoe J. (2017). A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. Dev. Cell 41, 243–261.e7. 10.1016/j.devcel.2017.04.002. [PubMed: 28457792]

Gunne-Braden A, Sullivan A, Gharibi B, Sheriff RSM, Maity A, Wang YF, Edwards A, Jiang M, Howell M, Goldstone R, et al. (2020). GATA3 mediates a fast, irreversible commitment to BMP4-driven differentiation in human embryonic stem cells. Cell Stem Cell 26, 693–706.e9. 10.1016/j.stem.2020.03.005. [PubMed: 32302522]

Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29. 10.1016/j.cell.2021.04.048. [PubMed: 34062119]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. 10.1016/j.molcel.2010.05.004. [PubMed: 20513432]

Henrique D, Abranches E, Verrier L, and Storey KG (2015). Neuromesodermal progenitors and the making of the spinal cord. Development 142, 2864–2875. 10.1242/dev.119768. [PubMed: 26329597]

Keller C, Arenkiel BR, Coffin CM, El-Bardeesy N, DePinho RA, and Capecchi MR (2004). Alveolar rhabdomyosarcomas in conditional Pax3:Fkhr mice: cooperativity of Ink4a/ARF and Trp53 loss of function. Genes Dev. 18, 2614–2626. [PubMed: 15489287]

Kent WJ, Zweig AS, Barber G, Hinrichs AS, and Karolchik D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26, 2204–2207. 10.1093/bioinformatics/btq351. [PubMed: 20639541]

Kinoshita M, Barber M, Mansfield W, Cui Y, Spindlow D, Stirparo GG, Dietmann S, Nichols J, and Smith A. (2021). Capture of mouse and human stem cells with features of formative pluripotency. Cell Stem Cell 28, 2180. 10.1016/j.stem.2021.11.002. [PubMed: 34861148]

Klemm SL, Shipony Z, and Greenleaf WJ (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 20, 207–220. 10.1038/s41576-018-0089-8. [PubMed: 30675018]

Knoepfler PS, Bergstrom DA, Uetsuki T, Dac-Korytko I, Sun YH, Wright WE, Tapscott SJ, and Kamps MP (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. Nucleic Acids Res. 27, 3752–3761. [PubMed: 10471746]

Knoepfler PS, Calvo KR, Chen H, Antonarakis SE, and Kamps MP (1997). Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. Proc. Natl. Acad. Sci. USA 94, 14553–14558. 10.1073/pnas.94.26.14553. [PubMed: 9405651]

Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, and Raychaudhuri S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296. 10.1038/s41592-019-0619-0. [PubMed: 31740819]

Kurup JT, Han Z, Jin W, and Kidder BL (2020). H4K20me3 methyltransferase SUV420H2 shapes the chromatin landscape of pluripotent embryonic stem cells. Development 147, dev188516. 10.1242/dev.188516.

Lang D, Brown CB, Milewski R, Jiang YQ, Lu MM, and Epstein JA (2003). Distinct enhancers regulate neural expression of Pax7. Genomics 82, 553–560. 10.1016/s0888-7543(03)00178-2. [PubMed: 14559212]

Lange C, Turrero Garcia M, Decimo I, Bifari F, Eelen G, Quaegebeur A, Boon R, Zhao H, Boeckx B, Chang J, et al. (2016). Relief of hypoxia by angiogenesis promotes neural stem cell differentiation by targeting glycolysis. EMBO J. 35, 924–941. 10.15252/embj.201592372. [PubMed: 26856890]

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. 10.1186/gb-2009-10-3-r25. [PubMed: 19261174]

Li H, and Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. 10.1093/bioinformatics/btp698. [PubMed: 20080505]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Lilja KC, Zhang N, Magli A, Gunduz V, Bowman CJ, Arpke RW, Darabi R, Kyba M, Perlingeiro R, and Dynlacht BD (2017). Pax7 remodels the chromatin landscape in skeletal muscle stem cells. PLoS One 12, e0176190. 10.1371/journal.pone.0176190. [PubMed: 28441415]

Liu L, Cheung TH, Charville GW, and Rando TA (2015). Isolation of skeletal muscle stem cells by fluorescence-activated cell sorting. Nat. Protoc. 10, 1612–1624. 10.1038/nprot.2015.110. [PubMed: 26401916]

Magli A, Baik J, Mills LJ, Kwak IY, Dillon BS, Mondragon Gonzalez R, Stafford DA, Swanson SA, Stewart R, Thomson JA, et al. (2019). Time-dependent Pax3-mediated chromatin remodeling and cooperation with Six4 and Tead2 specify the skeletal myogenic lineage in developing mesoderm. PLoS Biol. 17, e3000153. 10.1371/journal.pbio.3000153. [PubMed: 30807574]

Magli A, and Perlingeiro RRC (2017). Myogenic progenitor specification from pluripotent stem cells. Semin. Cell Dev. Biol. 72, 87–98. 10.1016/j.semcdb.2017.10.031. [PubMed: 29107681]

Mansouri A, and Gruss P. (1998). Pax3 and Pax7 are expressed in commissural neurons and restrict ventral neuronal identity in the spinal cord. Mech. Dev. 78, 171–178. 10.1016/s0925-4773(98)00168-3. [PubMed: 9858722]

Martello G, and Smith A. (2014). The nature of embryonic stem cells. Annu. Rev. Cell Dev. Biol. 30, 647–675. 10.1146/annurev-cellbio-100913-013116. [PubMed: 25288119]

McKinnell IW, Ishibashi J, Le Grand F, Punch VGJ, Addicks GC, Greenblatt JF, Dilworth FJ, and Rudnicki MA (2008). Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. Nat. Cell Biol. 10, 77–84. 10.1038/ncb1671. [PubMed: 18066051]

Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, and Sartorelli V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol. Cell 51, 606–617. 10.1016/j.molcel.2013.07.022. [PubMed: 23993744]

Oustanina S, Hause G, and Braun T. (2004). Pax7 directs postnatal renewal and propagation of myogenic satellite cells but not their specification. Embo J. 23, 3430–3439. [PubMed: 15282552]

Pera MF, and Rossant J. (2021). The exploration of pluripotency space: charting cell state transitions in peri-implantation development. Cell Stem Cell 28, 1896–1906. 10.1016/j.stem.2021.10.001. [PubMed: 34672948]

Pillai A, Mansouri A, Behringer R, Westphal H, and Goulding M. (2007). Lhx1 and Lhx5 maintain the inhibitory-neurotransmitter status of interneurons in the dorsal spinal cord. Development 134, 357–366. 10.1242/dev.02717. [PubMed: 17166926]

Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. 10.1093/bioinformatics/btq033. [PubMed: 20110278]

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, and Wysocka J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283. [PubMed: 21160473]

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, and Aiden EL (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680. 10.1016/j.cell.2014.11.021. [PubMed: 25497547]

Relaix F, Bencze M, Borok MJ, Der Vartanian A, Gattazzo F, Mademtzoglou D, Perez-Diaz S, Prola A, Reyes-Fernandez PC, Rotini A, and Taglietti, t. (2021). Perspectives on skeletal muscle stem cells. Nat. Commun. 12, 692. 10.1038/s41467-020-20760-6. [PubMed: 33514709]

Rodrigues AR, Yakushiji-Kaminatsui N, Atsuta Y, Andrey G, Schorderet P, Duboule D, and Tabin CJ (2017). Integration of Shh and Fgf signaling in controlling Hox gene expression in cultured limb cells. Proc. Natl. Acad. Sci. USA 114, 3139–3144. 10.1073/pnas.1620767114. [PubMed: 28270602]

Rojas A, De Val S, Heidt AB, Xu SM, Bristow J, and Black BL (2005). Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. Development 132, 3405–3417. 10.1242/dev.01913. [PubMed: 15987774]

Ryall JG, Cliff T, Dalton S, and Sartorelli V. (2015). Metabolic reprogramming of stem cell epigenetics. Cell Stem Cell 17, 651–662. 10.1016/j.stem.2015.11.012. [PubMed: 26637942]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Seale P, Sabourin LA, Girgis-Gabardo A, Mansouri A, Gruss P, and Rudnicki MA (2000). Pax7 is required for the specification of myogenic satellite cells. Cell 102, 777–786. [PubMed: 11030621]

Sharma A, Wasson LK, Willcox JA, Morton SU, Gorham JM, De-Laughter DM, Neyazi M, Schmid M, Agarwal R, Jang MY, et al. (2020). GATA6 mutations in hiPSCs inform mechanisms for maldevelopment of the heart, pancreas, and diaphragm. Elife 9, e53278. 10.7554/eLife.53278. [PubMed: 33054971]

Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, and Costantini F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev. Biol. 1, 4. 10.1186/1471-213x-1-4. [PubMed: 11299042]

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, and Satija R. (2019). Comprehensive integration of single-cell data. Cell 177, 1888–1902.e21. 10.1016/j.cell.2019.05.031. [PubMed: 31178118]

Stuart T, Srivastava A, Madad S, Lareau CA, and Satija R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods 18, 1333–1341. 10.1038/s41592-021-01282-5. [PubMed: 34725479]

Takemoto T, Uchikawa M, Kamachi Y, and Kondoh H. (2006). Convergence of Wnt and FGF signals in the genesis of posterior neural plate through activation of the Sox2 enhancer N-1. Development 133, 297–306. 10.1242/dev.02196. [PubMed: 16354715]

Tan G, and Lenhard B. (2016). TFBSTools: an R/bioconductor package for transcription factor binding site analysis. Bioinformatics 32, 1555–1556. 10.1093/bioinformatics/btw024. [PubMed: 26794315]

Tan G, and Lenhard B. (2016). TFBSTools: An R/bioconductor package for transcription factor binding site analysis. Bioinformatics 32, 1555–1556. 10.1093/bioinformatics/btw024. [PubMed: 26794315]

Tierney MT, and Sacco A. (2016). Satellite cell heterogeneity in skeletal muscle homeostasis. Trends Cell Biol. 26, 434–444. 10.1016/j.tcb.2016.02.004. [PubMed: 26948993]

Torres-Padilla ME, and Chambers I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development 141, 2173–2181. 10.1242/dev.102624. [PubMed: 24866112]

Trapnell C, Pachter L, and Salzberg SL (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. [PubMed: 19289445]

Vadasz S, Marquez J, Tulloch M, Shylo NA, and García-Castro MI (2013). Pax7 is regulated by cMyb during early neural crest development through a novel enhancer. Development 140, 3691–3702. 10.1242/dev.088328. [PubMed: 23942518]

Verdeguer F, Le Corre S, Fischer E, Callens C, Garbay S, Doyen A, Igarashi P, Terzi F, and Pontoglio M. (2010). A mitotic transcriptional switch in polycystic kidney disease. Nat. Med. 16, 106–110. 10.1038/nm.2068. [PubMed: 19966811]

Visel A, Rubin EM, and Pennacchio LA (2009). Genomic views of distant-acting enhancers. Nature 461, 199–205. 10.1038/nature08451. [PubMed: 19741700]

von Maltzahn J, Jones AE, Parks RJ, and Rudnicki MA (2013). Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. Proc. Natl. Acad. Sci. USA 110, 16474–16479. 10.1073/pnas.1307680110. [PubMed: 24065826]

Wamaitha SE, del Valle I, Cho LTY, Wei Y, Fogarty NME, Blakeley P, Sherwood RI, Ji H, and Niakan KK (2015). Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. Genes Dev. 29, 1239–1255. 10.1101/gad.257071.114. [PubMed: 26109048]

Wang X, Xiang Y, Yu Y, Wang R, Zhang Y, Xu Q, Sun H, Zhao ZA, Jiang X, Wang X, et al. (2021). Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. Cell Res. 31, 526–541. 10.1038/s41422-021-00477-x. [PubMed: 33608671]

Weidgang CE, Russell R, Tata PR, Kühl SJ, Illing A, Müller M, Lin Q, Brunner C, Boeckers TM, Bauer K, et al. (2013). TBX3 directs cell-fate decision toward mesendoderm. Stem Cell Rep. 1, 248–265. 10.1016/j.stemcr.2013.08.002.

Wen J, Hu Q, Li M, Wang S, Zhang L, Chen Y, and Li L. (2008). Pax6 directly modulate Sox2 expression in the neural progenitor cells. Neuroreport 19, 413–417. 10.1097/WNR.0b013e3282f64377. [PubMed: 18287938]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, and Grüning BA (2020). Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res. 48, W177–W184. 10.1093/nar/gkaa220. [PubMed: 32301980]

Wu J, Matthias N, Lo J, Ortiz-Vitali JL, Shieh AW, Wang SH, and Darabi R. (2018). A myogenic double-reporter human pluripotent stem cell line allows prospective isolation of skeletal muscle progenitors. Cell Rep. 25, 1966–1981.e4. 10.1016/j.celrep.2018.10.067. [PubMed: 30428361]

Xi H, Langerman J, Sabri S, Chien P, Young CS, Younesi S, Hicks M, Gonzalez K, Fujiwara W, Marzi J, et al. (2020). A human skeletal muscle atlas identifies the trajectories of stem and progenitor cells across development and from human pluripotent stem cells. Cell Stem Cell 27, 181–185. 10.1016/j.stem.2020.06.006. [PubMed: 32619514]

Yin H, Price F, and Rudnicki MA (2013). Satellite cells and the muscle stem cell niche. Physiol. Rev. 93, 23–67. 10.1152/physrev.00043.2011. [PubMed: 23303905]

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, and Smith A. (2008). The ground state of embryonic stem cell self-renewal. Nature 453, 519–523. 10.1038/nature06968. [PubMed: 18497825]

Young RA (2011). Control of the embryonic stem cell state. Cell 144, 940–954. 10.1016/j.cell.2011.01.032. [PubMed: 21414485]

Zhang N, Mendieta-Esteban J, Magli A, Lilja KC, Perlingeiro RCR, Marti-Renom MA, Tsirigos A, and Dynlacht BD (2020). Muscle progenitor specification and myogenic differentiation are associated with changes in chromatin topology. Nat. Commun. 11, 6222. 10.1038/s41467-020-19999-w. [PubMed: 33277476]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]

Zhou HY, Katsman Y, Dhaliwal NK, Davidson S, Macpherson NN, Sakthidevi M, Collura F, and Mitchell JA (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes Dev. 28, 2699–2711. 10.1101/gad.248526.114. [PubMed: 25512558]

Zhou W, Choi M, Margineantu D, Margaretha L, Hesson J, Cavanaugh C, Blau CA, Horwitz MS, Hockenbery D, Ware C, and Ruohola-Baker H. (2012). HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. EMBO J. 31, 2103–2116. 10.1038/emboj.2012.71. [PubMed: 22446391]

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, and Chanda SK (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. 10, 1523. 10.1038/s41467-019-09234-6. [PubMed: 30944313]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- Transcriptional and epigenetics characterization of ESCs exiting pluripotency
- Presomitic mesoderm gene programs and chromatin accessibility at single-cell resolution
- Single-cell multiomics of ESCs acquiring neurogenic and myogenic fates
- Identification of an enhancer directing Pax7 and neurogenic and myogenic gene programs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

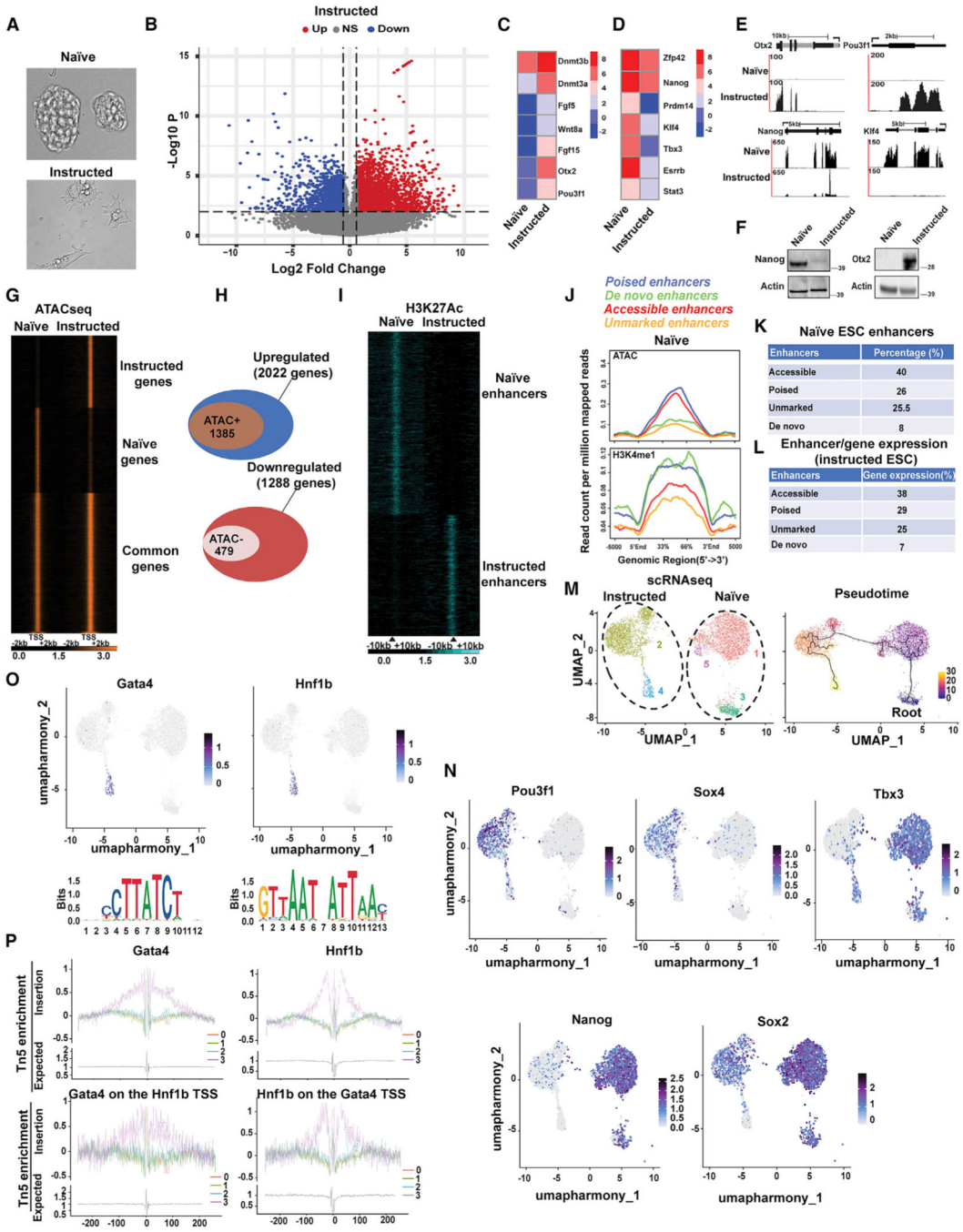


Figure 1. Bulk and single-cell analyses of transcriptomes, enhancers, and chromatin accessibility in naive and Bmp4-treated ESCs

(A) Morphology of ESCs cultured 2i+LIF conditions (naive) or exposed to Bmp4 in serum-free medium for 48 h (instructed).

(B) Transcriptome changes of instructed ESCs. Each point represents RPKM values obtained for a given transcript in RNA-seq analysis from instructed ESCs. Abscissa represents magnitude of log2 fold change, and ordinate indicates statistical significance (-log10 p value). NS, not significant.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(C and D) Heatmaps representing expression of selected transcripts in naive and instructed ESCs (log2 fold change).
(E) RNA-seq tracks of Otx2, Pou3f1, Nanog, and Klf4 in naive and instructed ESCs.
(F) Immunoblot of Nanog or Otx2 protein in naive and instructed ESCs. Beta-actin was employed as loading control.
(G) Heatmap representing ATAC-seq signal intensity in naive and instructed ESCs. Signal is centered −2/+2 Kb from the transcriptional start site (TSS). Peak calling was determined by MACS 1.4.2 with a p value of 1.0E–05.
(H) Venn diagrams intersecting upregulated genes acquiring ATAC-seq signal (top panel) and downregulated genes losing ATAC-seq signal (bottom panel) in instructed ESCs. Counts represent the number of individual genes identified by ANOVA test (fold change 1.5, minimum expression ≥ 1 RPKM) intersected with ATAC-seq^{+} TSSs in instructed ESCs (MACS 1.4.2, p = 1.0E–05).
(I) H3K27ac heatmap of H3K4me1^{+} regions in naive and instructed ESCs. Interval is −10/+10 Kb from the center of the peak signal.
(J) Averaged normalized tag intensities of ATAC-seq (top panel) or H3K4me1 (bottom panel) signal at genomic regions in naive ESCs acquiring H3K27ac in instructed ESCs.
(K) Percentage of genomic regions with distinct ATAC-seq and H3K4me1 enrichment in naive ESCs acquiring H3K27ac signal in instructed ESCs.
(L) Percentage of upregulated genes in instructed ESCs associated with genomic regions with distinct ATAC-seq and H3K4me1 enrichment in naive ESCs.
(M) scRNA-seq UMAP of naive and instructed ESCs (left panel). Pseudotime ordering of naive and instructed ESCs clusters (right panel). The heatmap scale represents units of progress, with “0” located at the root and “30” at the end of the trajectory.
(N) Expression of Pou3f1, Sox4, Tbx3, Nanog, and Sox2 transcripts in naive and instructed ESCs clusters.
(O) Expression and DNA motif enrichment for Gata4 and Hnfb1 in naive and instructed ESCs clusters.
(P) Overall footprinting of Gata4 (p = 0; percentage of occupied sites = 38.83) and Hnf1 (p = 4.897e–256; percentage of occupied sites = 27.92) in instructed ESCs (top panel). Gata4 footprinting at Hnf1b TSS (p = 1.49e–21; percentage of occupied sites = 30.76), and Hnfb1 footprinting at Gata4 TSS (p = 8.32e–30; percentage of occupied sites = 23.42) (bottom panel). The x axis in the footprinting represents nucleotides from the DNA-binding motif located at 0.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

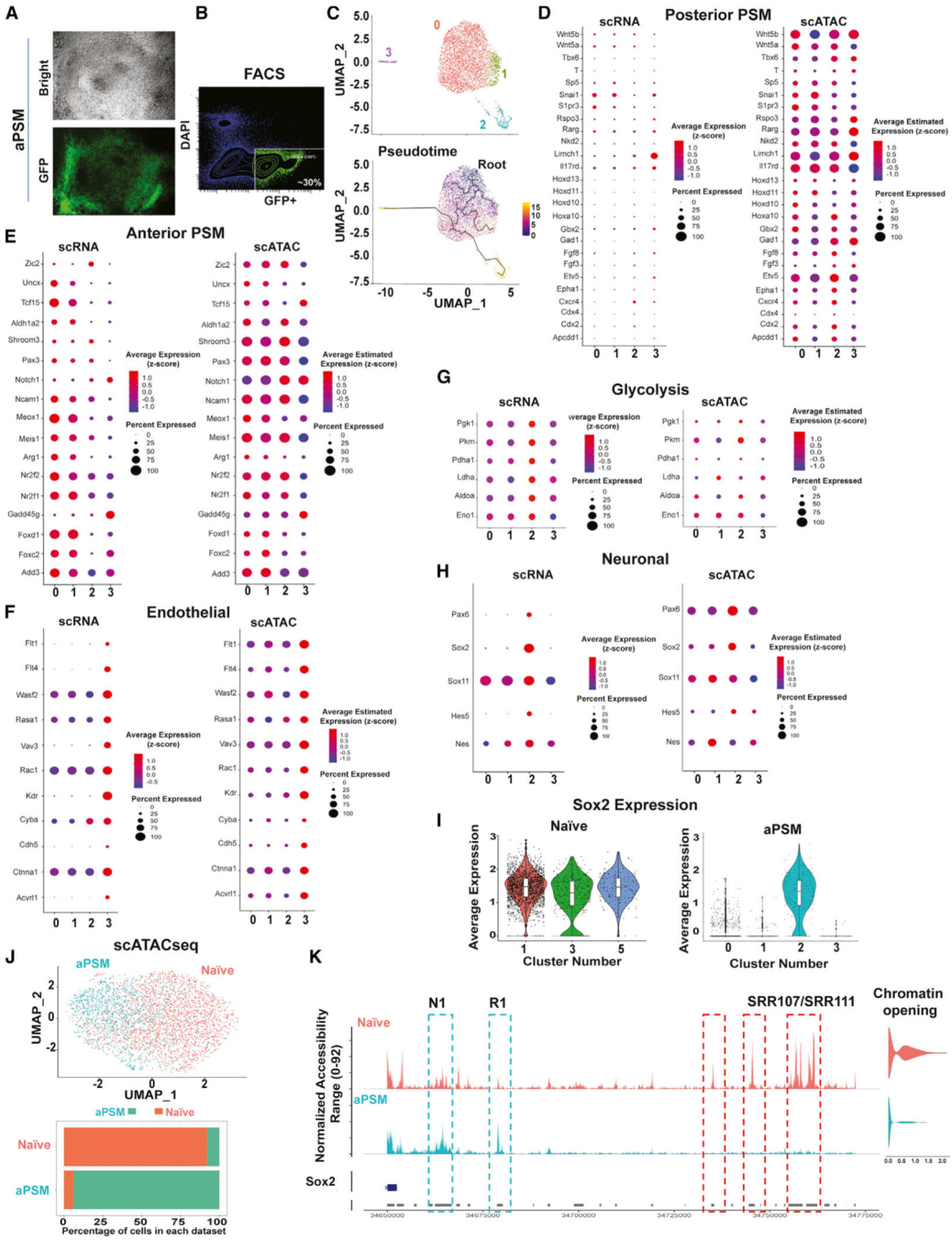


Figure 2. Single-cell analysis of transcriptome and chromatin accessibility of Pax3-GFP-positive ESCs

(A) Bright-field and fluorescence microscopy of ESCs Pax3-GFP cultured in conditions favoring acquisition of anterior presomitic mesoderm (aPSM) fate.

(B) FACS isolation of GFP$^+$ ESCs Pax3-GFP.

(C) scRNA-seq UMAP graph of FACS-isolated GFP$^+$ ESCs Pax3-GFP (top panel) and pseudotime ordering (bottom panel). The pseudotime heatmap scale represents units of progress, with 0 located at the root of the trajectory.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(D–H) scRNA-seq and scATAC-seq Louvain dot plots of selected (D) posterior PSM markers, (E) aPSM markers, (F) endothelial markers, (G) glycolytic genes, and (H) neuronal markers.

(I) Violin plots representing Sox2 expression in naive and GFP^+ ESC Pax3-GFP (aPSM).

(J) scATAC-seq clustering of naive and GFP^+ ESC Pax3-GFP (aPSM) (top panel), and cell percentage of each cluster (bottom panel).

(K) scATAC-seq tracks of the Sox2 locus in naive and GFP^+ ESC Pax3-GFP (aPSM). SRR107/SRR111 indicate Sox2 enhancers active in naive ESCs. N1 denotes a Sox2 enhancer active in neurogenic cells. Next to N1, a genomic region (R1) with increased chromatin accessibility in aPSM cells. Violin plots indicate quantification of Sox2 chromatin accessibility (opening) in naive and aPSM cells, respectively.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

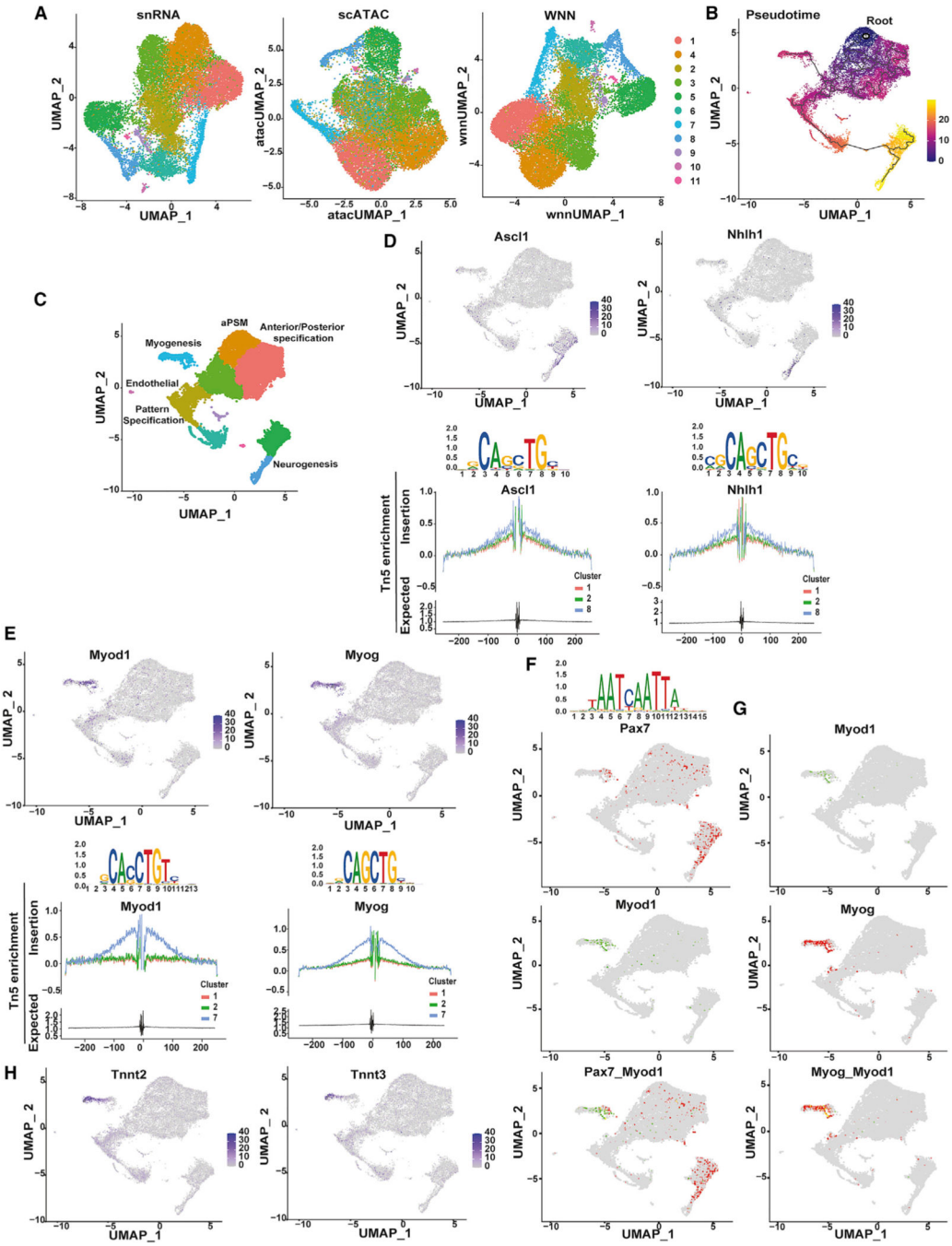


Figure 3. Single-cell omics of ESC-derived cell lineages

(A) HIFLR cells clustering based on snRNA, scATAC, and WNN analysis performed on FACS (GFP^+)-isolated cells. The same colors were employed to identify clusters in snRNA-seq and scATAC-seq.

(B) Clusters' trajectory inference (pseudotime). The heatmap represents units of progress, with 0 located at the root of the trajectory.

(C) Clustering derived from trajectory inference.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(D) Expression, DNA binding motifs, and footprinting (from top to bottom) for Ascl1 and Nhlh1. The x axis in the footprinting represents nucleotides from the DNA binding motif located at 0.

(E) Expression, DNA-binding motifs, and footprinting (from top to bottom) for Myod1 and Myogenin.

(F) Pax7 DNA-binding motif (top panel) and paired-plots expression of Pax7 and Myod1 (bottom panel).

(G) Paired-plots expression of Myod1 and Myogenin.

(H) Tnnt2 and Tnnt3 expression.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

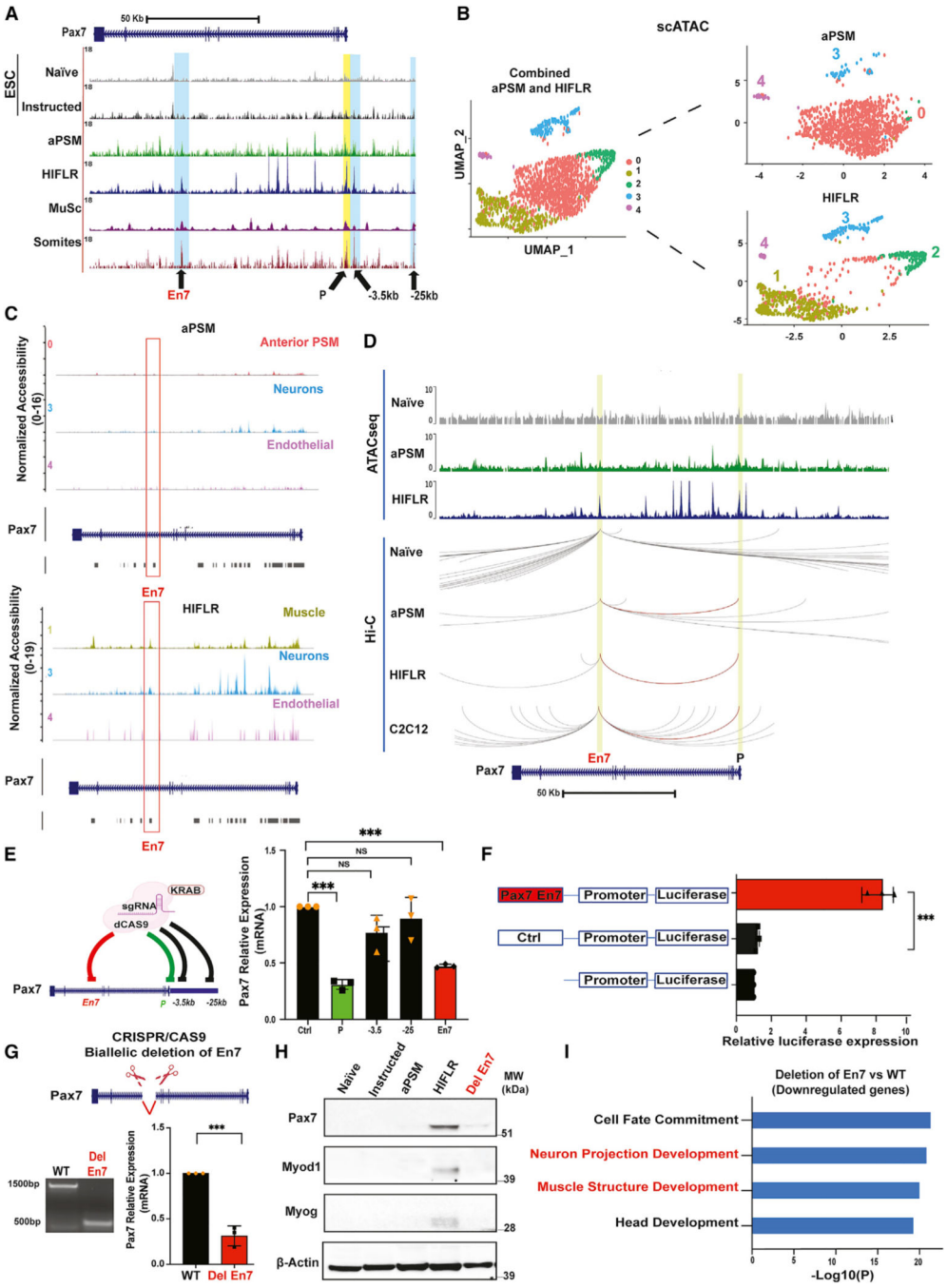


Figure 4. Identification and characterization of genomic regions regulating Pax7 expression

(A) Genome browser representation of ATAC-seq tracks at the Pax7 locus in naive, instructed ESCs, aPSM, HIFLR, MuSCs, and E12.5 YFP^{+} dissected somites from Pax7^{+}/Cre; Rosa26-YFP mice. En7 indicates a region located within the Pax7 seventh intron; P, promoter; −3.5 and −25 Kb, regions upstream the Pax7 TSS.

(B) scATAC-seq clusters of aPSM and HIFLR cells (left panel, combined clusters; right panel, aPSM and HIFLR individual clusters).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(C) scATAC-seq tracks at the Pax7 locus in aPSM (top panel) and HIFLR (bottom panel) clusters.
(D) ATAC-seq tracks and Hi-C interactions at the Pax7 locus in naive ESCs, aPSM, HIFLR, and C2C12 cells. The red line indicates En7-promoter interactions.
(E) Scheme representing gRNA-mediated dCas9-KRAB targeting of the indicated Pax7 regions in myogenic C2C12 cells (left panel). Quantitative PCR was employed to measure Pax7 mRNA in myogenic C2C12 cells transfected with dCas9-KRAB and specific or control gRNAs (right panel) Data are represented as mean ±SD (n = 3). Significance is displayed as ***$p < 0.001$.
(F) Luciferase assay in C2C12 cells transfected with the indicated reporter constructs. Data are represented as mean ±SD (n = 3). Significance is displayed as ***$p < 0.001$.
(G) Scheme representing biallelic En7 deletion (top panel). Genomic DNA electrophoresis documenting biallelic En7 deletion. Quantitative PCR was employed to measure Pax7 mRNA in control and En7-deleted HIFLR cells (bottom panel). Data are represented as mean ±SD (n = 3). Significance is displayed as ***$p < 0.001$.
(H) Immunoblot of Pax7, MyoD, Myogenin, and beta-actin in naive, instructed, aPSM, HIFLR, and HIFLR DelEn7cells.
(I) Gene Ontology terms of downregulated transcripts in En7-deleted HIFLR cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cell Rep. Author manuscript; available in PMC 2022 November 09.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Histone H3 (acetyl K27) - ChIP Grade	Abcam	ab4729, Lot: GR3374555-1
Anti-Histone H3 (mono methyl K4)	Abcam	ab8895, Lot: GR3407156-1
Anti-OTX2 (D7Y3J)	Cell Signaling Technology	mAb #11943, Lot#1
Anti-Nanog (A-11)	Santa Cruz Biotechnology	sc-374001, Lot#E0521
Anti-PAX7	DSHB	N/A
β-Actin (N-21)	Santa Cruz Biotechnology	sc-130656, Lot #C2515
MyoD (4H207)	Santa Cruz Biotechnology	sc-71629, Lot #G2012
MyoG (F5D)	Santa Cruz Biotechnology	sc-12732, Lot #A0620
Goat Anti Rabbit IgG-HRP	Azure biosystem	AC2114, Lot 210812-57
Goat Anti mouse IgG-HRP	Azure biosystem	AC2115, Lot 210204-50
Bacterial and virus strains		
NEB® 5-alpha Competent E. coli (High Efficiency)	NEB	C2987H
One Shot™ TOP10 Chemically Competent E. coli	ThermoFisher scientific	C404010
Chemicals, peptides, and recombinant proteins		
Knock-Out™ Serum Replacement	ThermoFisher	10828028
Recombinant Protein-Rspo3	R&D Biosystems	4120-RS-025
Leukemia inhibitory factor-LIF	Millipore	ESG1106
Recombinant Murine BMP4	Peprotech	315–27
BSA solution	Gibco	101020–021
LDN193189	Stemgent	04–0074
CHIRON99021	Stemgent	04-0004-02
Recombinant Murine FGF	Peprotech	450-33
Recombinant Murine IGF	Peprotech	250-19
Recombinant Mouse HGF	R&D Biosystems	2207-HG-2-025
Dynabeads Protein G	ThermoFisher Scientific	10004D
TRIzol Reagent	Thermo Fisher Scientific	15596018

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Polybrene	Sigma	TR-1003-G
Protease Inhibitor	Roche	11836170001
Phase Lock Gel Heavy	5 PRIME	2302810
Chloroform	Sigma	67-66-3
10% SDS	Invitrogen	15553-027
Neurobasal Medium	Gibco	21103-049
DMEM/F12 (1:1)	Gibco	11320-033
Dimethyl sulfoxide	Sigma	276855
2-Mercaptoethanol	Sigma	M6250
Dulbecco's Modified Eagle Medium (DMEM)	Gibco	12439-054
Pen Strep Glutamine (100x)	Gibco	10378-016
GlutaMAX (100x)	Gibco	35050-061
MEM NEAA	Gibco	11140-050
Sodium Pyruvate (100mM)	Gibco	11360-070
BSA Fraction V (7.5%)	Gibco	15260-037
Critical commercial assays		
DNeasy Blood & Tissue Kits	Qiagen	69504
QIAprep Spin Miniprep Kit	Qiagen	27106
Lenti-X™ Concentrator	Clontech	631231
Lipofectamine 2000	Invitrogen	11668-019
Power SYBR Green PCR Master Mix	Thermo Fisher Scientific	4368708
PCR purification kit	Qiagen	28106
qScript™ cDNA Synthesis Kit	Quantabio	95047-100
Arima HiC kit	Arima	A510008GFP
Accel-NGSO 2S Plus DNA	SWIFT Bioscience	A160140 v00
NEBNext Ultra II DNA library prep kit	NEB	#E7645
NEBNext Ultra II RNA library preparation kit for Illumina	NEB	#E7490
Chromium Single Cell 3' Library & Gel Bead Kit v3.1	10x Genomics	P/N 1000121
Deposited data		

Cell Rep. Author manuscript; available in PMC 2022 November 09.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RNA-seq, ATAC-seq, H327ac and H3K4me1 ChIP-seq, HiC, Single cell RNA-seq, Single cell ATAC-seq, Multiome single nuclei ATAC and gene expression	This paper	GSE198730
Experimental models: Cell lines		
HEK 293T cells	ATCC	CRL-1573
C2C12 cells	ATCC	GSC-6001G
ESC Pax3-GFP	Chal et al., 2015	N/A
ESC Pax3-GFP Del −3.5kb	This paper	N/A
ESC Pax3-GFP Del −3.5kb and −25kb	This paper	N/A
ESC Pax3-GFP Del En7	This paper	N/A
Oligonucleotides		
sgRNA sequences	Table S4	N/A
RT q-PCR oligos	Table S4	N/A
PCR oligos	Table S4	N/A
Recombinant DNA		
pSpCas9(BB)-2A-GFP (PX458)	Addgene	Plasmid #48138
pSpCas9(BB)-2A-GFP-gRNA1_−3.5kb	This paper	N/A
pSpCas9(BB)-2A-GFP-gRNA4_−3.5kb	This paper	N/A
pSpCas9(BB)-2A-GFP-gRNA1_−25kb	This paper	N/A
pSpCas9(BB)-2A-GFP-gRNA4_−25 kb	This paper	N/A
pSpCas9(BB)-2A-GFP-gRNA2_En7	This paper	N/A
pSpCas9(BB)-2A-GFP-gRNA4_En7	This paper	N/A
pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP	Addgene	Plasmid #71237
pLV hU6-sgRNA_Pax7_promoter_ hUbC-dCas9-KRAB-T2a-GFP	This paper	N/A
pLV hU6-sgRNA_Pax7_−3.5kb_ hUbC-dCas9-KRAB-T2a-GFP	This paper	N/A
pLV hU6-sgRNA_Pax7_−25kb_ hUbC-dCas9-KRAB-T2a-GFP	This paper	N/A
pLV hU6-sgRNA_Pax7_ En7_ hUbC-dCas9-KRAB-T2a-GFP	This paper	N/A
psPAX2	Addgene	Plasmid ##12260
pMD2.G	Addgene	Plasmid #12259
pGL4.26[luc2/minP/Hygro]	Promega	#E8441

Cell Rep. Author manuscript; available in PMC 2022 November 09.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pGL4.26-Ctrl	This paper	N/A
pGL4.26-En7	This paper	N/A
Software and algorithms		
BioRender		biorender.com
Prism Software, version 8	Graphpad Software	https://www.Graphpad.com
Metascape	Zhou et al. (2019)	https://metascape.org/gp/index.html#/main/step1
CellRanger ver. 3.1.0	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
Cell Ranger ATAC v1.1.0	10x Genomics	https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac
R	The R Project for Statistical Computing	https://www.r-project.org/
Seurat v4.1.0	Stuart et al., 2019	https://cran.r-project.org/web/packages/Seurat/index.html
Signac v1.5.0	Stuart et al., 2021	https://satijalab.org/signac
Harmony	Korsunsky et al., 2019	https://github.com/immunogenomics/harmony
Monocle3	Cao et al., 2019	https://cole-trapnell-lab.github.io/monocle3/
JASPAR 2020	Fornes et al. (2020)	https://bioconductor.org/packages/release/data/annotation/html/JASPAR2020.html
FBSTools	Tan and Lenhard (2016)	https://bioconductor.org/packages/release/bioc/html/TFBSTools.html
cellranger-arc version 2.0.0	10x Genomics	https://support.10xgenomics.com/single-cell-multiome-atac-gex/software/pipelines/latest/what-is-cell-ranger-arc
Seurat ver. 4.1.0	Hao et al., 2021	https://satijalab.org/seurat/

Cell Rep. Author manuscript; available in PMC 2022 November 09.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Signac ver.1.5.0	Stuart et al., 2021	https://satijalab.org/signac/news/index.html
MACS 1.4.2	Zhang et al., 2008	https://libraries.io/pypi/MACS
Bowtie/1.1.1	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
BedGraphToBigWig	Kent et al., 2010	https://www.encodeproject.org/software/bedgraphtobigwig/
Bedtools/2.29.2	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/content/installation.html
HOMER/4.11.1	Heinz et al., 2010	http://homer.ucsd.edu/homer/
Partek Genomics Suite 7.18	Partek Inc.	https://www.partek.com/partek-genomics-suite/
TopHat 2.1.1	Trapnell et al., 2009	https://ccb.jhu.edu/software/tophat/index.shtml
Trim Galore (version 0.6.6)	N/A	https://github.com/FelixKrueger/TrimGalore
Bcl2fastq/2.20.0	Illumina	https://emea.support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html
Juicer 1.6	Durand et al., 2016	https://github.com/aidenlab/juicer/releases
HiCExplorer3.6	Wolff et al., 2020	https://hicexplorer.readthedocs.io/en/latest/
HOMER	Heinz et al., 2010	http://homer.ucsd.edu/homer/
Custom scripts	This paper	https://doi.org/10.5281/zenodo.6889278