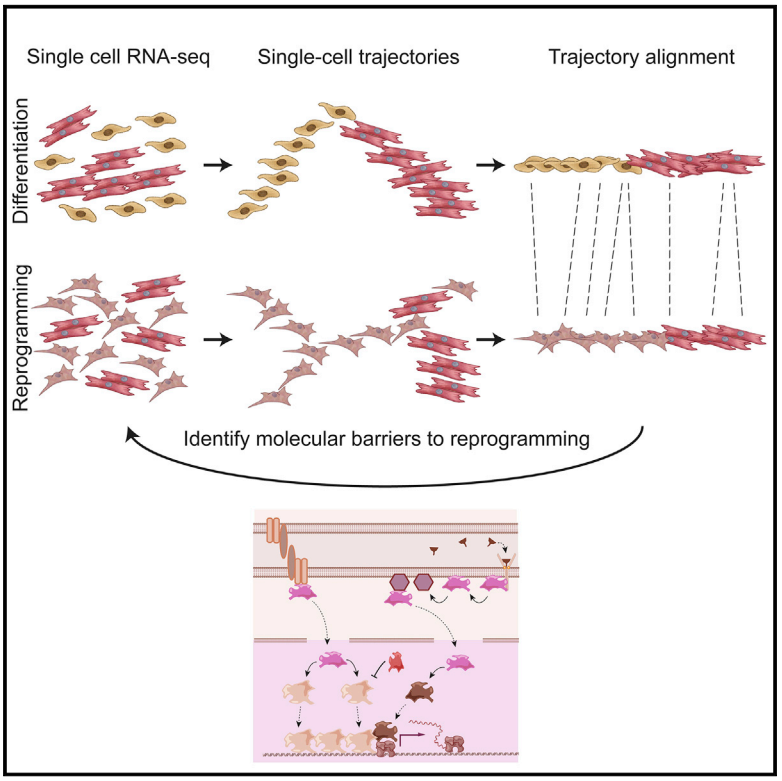
Article

Cell Systems

Aligning Single-Cell Developmental and Reprogramming Trajectories Identifies Molecular Determinants of Myogenic Reprogramming Outcome

Graphical Abstract



Authors

Davide Cacchiarelli, Xiaojie Qiu, Sanjay Srivatsan, ..., Tarjei S. Mikkelsen, John L. Rinn, Cole Trapnell

Correspondence

d.cacchiarelli@tigem.it (D.C.), coletrap@uw.edu (C.T.)

In Brief

Cellular reprogramming converts only a fraction of cells to the desired state. We analyze reprogramming of human fibroblasts to myotubes at pseudotemporal resolution using single-cell RNA-seq. We identified defects in BMP and insulin signaling as culprits using trajectory alignment, which enables quantitative comparison of gene expression kinetics across two biological processes.

Highlights

- Single-cell RNA-seq of myogenic reprogramming reveals barriers to conversion by MYOD
- Reprogramming and differentiation can be aligned by dynamic time warping
- Modulating insulin and BMP signaling greatly enhances myogenic conversion

Cacchiarelli et al., 2018, Cell Systems 7, 258–268
September 26, 2018
https://doi.org/10.1016/j.cels.2018.07.006

CellPress

OPEN ACCESS
CellPress

Cell Systems
Article

Aligning Single-Cell Developmental and Reprogramming Trajectories Identifies Molecular Determinants of Myogenic Reprogramming Outcome

Davide Cacchiarelli,[1,2,3,10,*] Xiaojie Qiu,[4,5] Sanjay Srivatsan,[5] Anna Manfredi,[1] Michael Ziller,[6] Eliah Overbey,[5] Antonio Grimaldi,[1] Jonna Grimsby,[3] Prapti Pokharel,[3] Kenneth J. Livak,[7,11] Shuqiang Li,[3] Alexander Meissner,[3,8,12] Tarjei S. Mikkelsen,[3,8,9,13] John L. Rinn,[3,8,14] and Cole Trapnell[4,5,10,15,*]

[1]Telethon Institute of Genetics and Medicine (TIGEM), Armenise/Harvard Laboratory of Integrative Genomics, Pozzuoli, Italy
[2]Department of Translational Medicine, University of Naples Federico II, Naples, Italy
[3]The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[4]Molecular & Cellular Biology Program, University of Washington, Seattle, WA, USA
[5]Department of Genome Sciences, University of Washington, Seattle, WA, USA
[6]Max-Planck Institute of Psychiatry, Munich, Germany
[7]Fluidigm Corporation, South San Francisco, CA, USA
[8]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA
[9]Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA
[10]These authors contributed equally
[11]Present address: Dana-Farber Cancer Institute, Boston, MA, USA
[12]Present address: Max Planck Institute for Molecular Genetics, Berlin, Germany
[13]Present address: 10x Genomics, Inc., Pleasanton, CA, USA
[14]Present address: Department of Biochemistry, BioFrontiers Institute, University of Colorado, Boulder, Boulder, CO, USA
[15]Lead Contact
*Correspondence: d.cacchiarelli@tigem.it (D.C.), coletrap@uw.edu (C.T.)
https://doi.org/10.1016/j.cels.2018.07.006

SUMMARY

Cellular reprogramming through manipulation of defined factors holds great promise for large-scale production of cell types needed for use in therapy and for revealing principles of gene regulation. However, most reprogramming systems are inefficient, converting only a fraction of cells to the desired state. Here, we analyze MYOD-mediated reprogramming of human fibroblasts to myotubes, a well-characterized model system for direct conversion by defined factors, at pseudotemporal resolution using single-cell RNA-seq. To expose barriers to efficient conversion, we introduce a novel analytic technique, trajectory alignment, which enables quantitative comparison of gene expression kinetics across two biological processes. Reprogrammed cells navigate a trajectory with branch points that correspond to two alternative decision points, with cells that select incorrect branches terminating at aberrant or incomplete reprogramming outcomes. Analysis of these branch points revealed insulin and BMP signaling as crucial molecular determinants of reprogramming. Single-cell trajectory alignment enables rigorous quantitative comparisons between biological trajectories found in diverse processes in development, reprogramming, and other contexts.

INTRODUCTION

During development, cells undergo drastic shifts in gene expression and epigenetic configuration as they pass from progenitor or stem cell states to their differentiated states in the adult organism. Nevertheless, developmental decisions can be "unmade" by ectopic expression of a small number of regulatory genes, as first shown by Davis et al., who converted murine fibroblasts to myotube-like cells by overexpressing MyoD, a key myogenic transcription factor (Davis et al., 1987). Numerous other reprogramming factors have since been discovered, most famously including four factors that reprogram many different cell types into induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006). Following the discovery of MyoD-mediated transdifferentiation in fibroblasts, Weintraub et al. explored MyoD's ability to convert a diverse set of starting cell types to myotubes (Weintraub et al., 1989). Interestingly, the ectopic expression of MyoD induced the expression of muscle-specific structural proteins without altering cell-specific functions. These studies, concomitant with contemporary heterokaryon experiments (Blau et al., 1983), indicated that neither intracellular transcription factors nor ectopically transduced MyoD were sufficient for complete reprogramming to the myogenic cell fate. Despite the following decades of research on cellular conversions, including extensive studies of myogenic reprogramming, the molecular determinants that impede or mediate reprogramming remain poorly characterized.

Reprogramming can generate dramatically heterogeneous cell populations, with cells reaching drastically different molecular outcomes. Single-cell genomics assays provide a means of

identifying the different outcomes and could reveal mechanisms that drive a cell to a particular one. For example, Truetlein et al. demonstrated through a single-cell transcriptome analysis that the overexpression of Ascl1, Brn2, and Myt1l generates both induced neurons and an alternative myocyte-like cell fate (Treutlein et al., 2016). However, synthesizing measurements from these diverse outcomes into a useful picture of the molecular mechanisms that determine cell fate remains extremely challenging. Recently, we developed an algorithm that can automatically reconstruct the sequence of expression changes executed by a cell undergoing differentiation or reprogramming from single-cell RNA sequencing (RNA-seq) data. Our algorithm, called Monocle, introduced the notion of "pseudotime," which measures each cell's progress through a biological process without the need for a priori knowledge of genes that define progression (Trapnell et al., 2014). Moreover, Monocle can pinpoint branches that lead a cell to alternative outcomes, which can reveal the genes that direct a cell to its ultimate fate. Monocle's unsupervised learning algorithm enables the discovery of key steps, roadblocks, and intermediate cellular states on the path to differentiation.

More recently, we released Monocle 2, which deploys a new algorithm framework called "reversed graph embedding" that dramatically improves the accuracy, scalability, and robustness of Monocle's trajectory inference (Qiu et al., 2017a, 2017b). In particular, Monocle 2 was far better at identifying branch points that correspond to fate decisions made during cell differentiation. Despite the improvements in Monocle 2, the toolkit is still largely focused on the analysis of a single biological process. However, many experiments call for the comparison of two developmental processes. For example, one may wish to compare the impact of a particular mutation on the development of a given cell type relative to wild-type cells. Comparing two pseudotime trajectories is not straightforward because there is no universal "unit" of pseudotime. A new algorithm is needed to match up the most similar segments of each trajectory so that any differences that remain between them can be identified.

Here, we develop a novel approach for aligning two pseudotime trajectories and use it to compare human fibroblasts undergoing MYOD-mediated myogenic reprogramming to normal muscle cell development. Through an unsupervised single-cell RNA-seq analysis, we reveal that the trajectory includes branch points corresponding to key "checkpoints" in the process. Cells that travel down the correct branch progress, while those that travel down the alternatives fail to convert to myotubes. We show that the myogenic conversion trajectory does contain a path that is similar to normal myoblast differentiation but that many cells diverge from this path toward unproductive reprogramming outcomes.

RESULTS

Pseudotemporal Analysis of MYOD Overexpression Reveals Multiple Reprogramming Outcomes

We sought to compare the reprogramming activation of the skeletal muscle genes in non-muscle cells to their developmentally regulated onset in normal myoblasts. Previously, we analyzed primary human skeletal muscle myoblast (HSMM) differentiation using single-cell RNA-seq by culturing HSMMs to optimal confluence and then withdrawing mitogens from the growth medium (GM) (Trapnell et al., 2014). Within 24 hr of the switch, HSMMs had exited the cell cycle, aligned with one another, and by 72 hr in differentiation (DM) had formed myosin heavy-chain (MHC)-positive myotubes. Monocle showed that HSMM cells follow a largely linear trajectory, albeit with a single small branch corresponding to non-differentiating myoblasts (i.e., "reserve cells") (Qiu et al., 2017a, 2017b). As cells followed this trajectory, they downregulated markers of the active proliferation (e.g., *CCNB2*) and then activated transcriptional regulators of the skeletal muscle program (e.g., *MYOG*), followed by genes associated with contractile myotubes (*MYH3*, *DMD*, *ENO3*, and others). To compare myoblast differentiation to myogenic reprogramming, we needed to engineer a fibroblast cell line in which the myoblast expression program could be induced in a tightly controlled manner.

To analyze myogenic reprogramming at single-cell resolution, we first derived a doxycycline-inducible MYOD fibroblast line (hFib-MyoD), enabling us to synchronize the onset of MYOD expression. The hFib-MyoD line also harbors a constitutively expressed human telomerase gene to alleviate passage-dependent stress, which we previously showed drastically improves uniformity and synchrony of induced pluripotency across cells (Cacchiarelli et al., 2015). 72 hr of *MYOD* transgene expression was sufficient to produce conversion of fibroblasts to MHC-positive cells. However, as reported by previous studies (Salvatori et al., 1995), myogenic reprogramming of human fibroblasts is a very inefficient process. Only a small proportion of cells are successfully reprogrammed, and the result is small, mononucleated, myocyte-like cells (Figures 1A and 1B) rather than long, multinucleated myotubes as generated by HSMM. The number of reprogrammed cells did not increase with longer time courses (not shown).

To investigate the molecular basis driving HSMM and hFib-MyoD through the process of myogenic identity, we sampled single cells every 24 hr post myogenic induction and performed deep, single-cell, full-length RNA-seq using the Fluidigm C1 platform. We recovered a total of 466 high-quality libraries from two independent reprogramming time series experiments. Although some markers of myotube formation, such as *ENO3*, were induced by MYOD, others including *DMD* and *MYOG* showed expression patterns inconsistent with HSMM (Figure 1C). Few hFib-MyoD cells expressed MYOG even at 72 hr, which is required for the upregulation of many genes needed for terminal differentiation and contraction, including *DMD* (Bentzinger et al., 2012). *CDK1* and other genes associated with active proliferation, which were rapidly downregulated in HSMM within 24 hr following the switch to the differentiation medium, were more gradually lost, with some cells still expressing them even at 72 hr. The delayed cell-cycle exit may explain why hFib-MyoD wells contained greater than 3-fold cells than HSMM wells despite being seeded at comparable initial density (Figure 1B). Comparing average expression levels in cells collected at each time point revealed that few of the 653 genes that were significantly differentially expressed as a function of time in myoblasts were regulated to the same extent in hFib-MyoD (Figure 1D).

Single-Cell Trajectory Branch Points Correspond to Reprogramming Barriers

We next sought to identify the subset of hFib-MyoD cells that reached a muscle-like expression program and define the

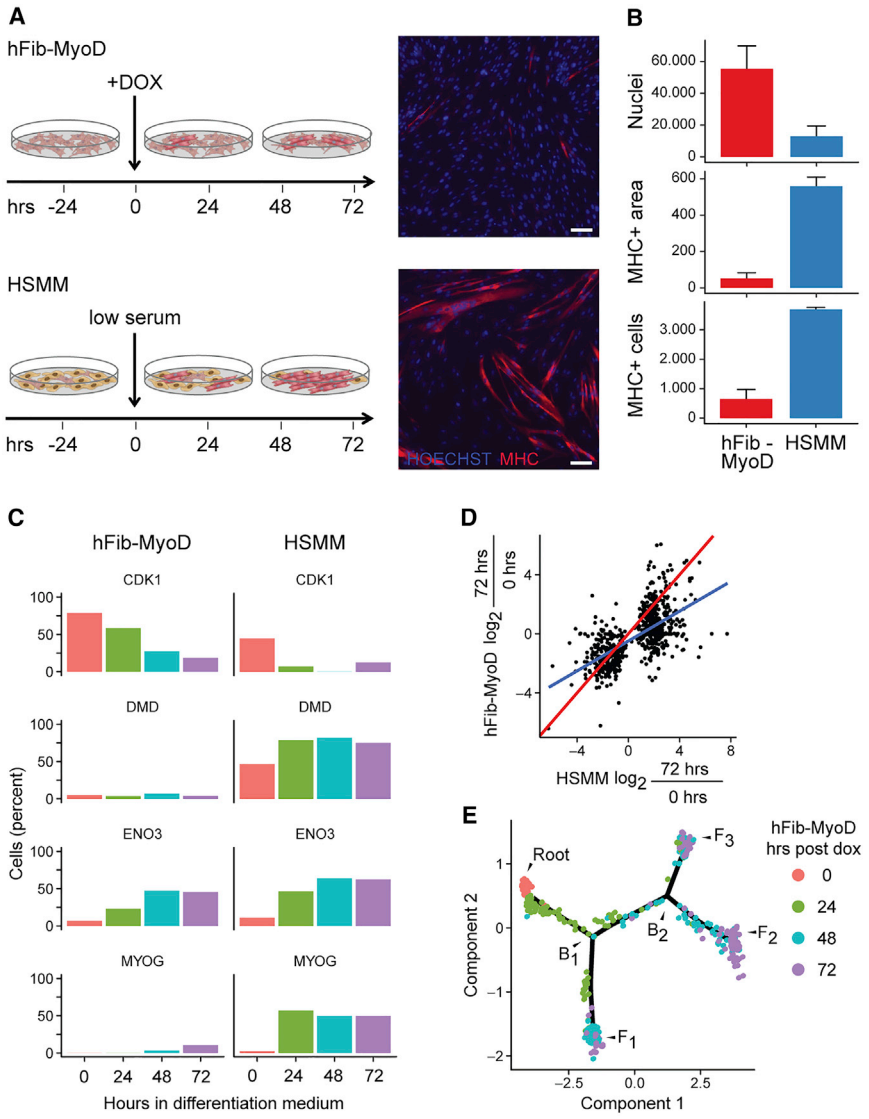


Figure 1. hFib-MyoD Is a Human Fibroblast Line that Converts to Myotube-like Cells upon Doxycycline Induction

(A) Immunostaining of hFib-MyoD and HSMM with muscle-specific anti-myosin-heavy chain (MHC) antibodies 72 hr post induction of MYOD-mediated reprogramming with doxycycline (dox) or myoblast differentiation via serum switch, respectively. Scale bar represents 100 μm.

(B) Counts of MHC+ cells, size of MHC+ cells in pixels, and total nuclei as measured by automated image processing scripts (see STAR Methods). Error bars indicate SD across replicates.

(C) Fraction of cells in which mRNAs for selected muscle markers were detected via full-length single-cell RNA-seq.

(D) Fold changes in average expression level of genes significantly differentially expressed (FDR < 5%) between 0 and 72 hr in differentiating myoblasts compared to their corresponding changes in hFib-MyoD. The blue line indicates a linear regression, while the red line illustrates perfect concordance.

(E) The single-cell trajectory reconstructed by Monocle 2 for hFib-MyoD cells undergoing myogenic reprogramming. Cells start at the root and progress to one of three alternative reprogramming outcomes, denoted by F_1, F_2, and F_3. To reach these fates, cells must pass through branch point B_1. Cells that do not proceed to F_1 must then choose between F_2 or F_3 at branch point B_2.

stepwise changes in gene expression that lead to conversion failure. Since both differentiation and reprogramming are characterized by high levels of asynchrony, obscuring the sequence of expression changes in such processes, we applied a recently improved version of Monocle (Qiu et al., 2017a) to hFib-MyoD. Monocle 2 reconstructed a trajectory capturing the progression of single cells through myogenic reprogramming (Figure 1E), which contained three termini (denoted "F_1," "F_2," and "F_3") corresponding to three distinct reprogramming outcomes. Cells from two independent biological replicates were found at each of the three outcomes (Figure S1). To reach these outcomes, cells passed through at least one of two branch points (denoted "B_1" and "B_2"). The trajectory's root was populated largely by cells collected at the beginning of the experiment, while the three termini of the tree were populated by cells collected following the switch to doxycycline-containing media.

We next assessed the expression of genes regulated in myoblast differentiation in cells at the three different outcomes of the hFib-MyoD trajectory. The cells at outcome F_1 showed expression of genes upregulated early in myoblast differentiation, such as *ENO3*, but maintained strong expression of genes needed for active proliferation such as *CDK1*. (Figure 2A) We next examined the pseudotemporal kinetics of *CDK1*, *ENO3*, and *TNNT2* as cells traveled from the root through branch point B_1 and then either to outcome F_1 or toward a second branch point B_2. While *ENO3* and *TNNT2* were upregulated to a similar extent on both paths, *CDK1* levels dropped markedly in cells traveling to the branch B_2 but remained high in cells traveling to F_1 (Figure 2B). A global differential analysis comparing the two paths away from branch point B_1 revealed 173 genes with branch-dependent expression (false discovery rate [FDR] < 1%), most of which were associated with the cell cycle. (Figure 2C; Table S1). We examined our previously collected HSMM data and found that in contrast to hFib-MyoD, key genes that mark proliferating cells (*CCNB2* and *CDK1*) are almost never co-expressed with early myogenic markers (e.g., *ENO3* and *TNNT2*) in these cells (Figure 3). Although hFib-MyoD cells at F_1 co-expressed numerous markers of both proliferation and the early myoblast differentiation program, they expressed few markers of mature, contractile myotubes. In normal differentiating myoblasts, proliferation genes are downregulated prior to the onset of skeletal muscle genes, indicating that F_1 may be an aberrant cellular state within the myogenic reprogramming process.

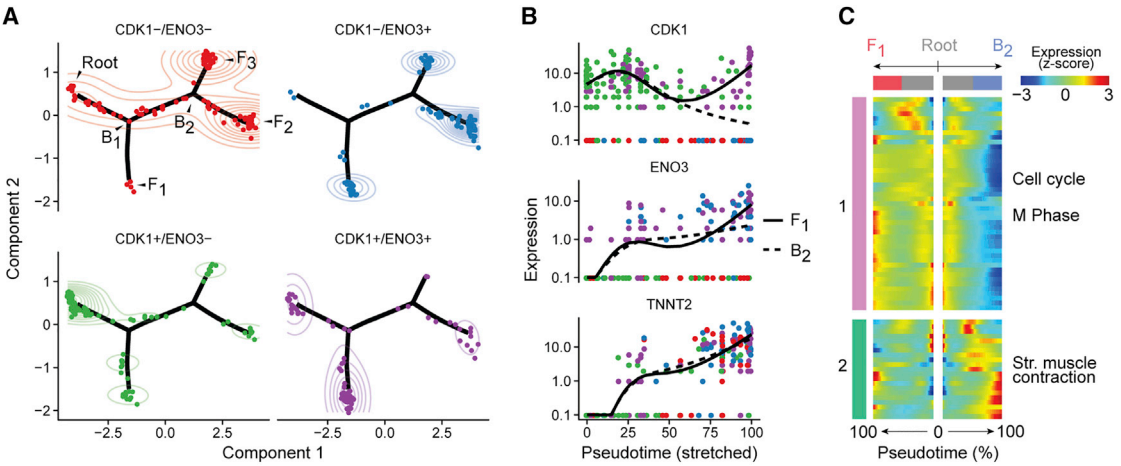


Figure 2. Single-Cell Trajectory Analysis with Monocle 2 Identifies Three MYOD-Mediated Myogenic Reprogramming Outcomes in hFib-MyoD

(A) Cells along the trajectory divided into four groupings based on expression of CDK1 and ENO3. Contour plots indicate two-dimensional Gaussian kernel density estimates. Cells are colored according to the facet in which they reside. hFib-MyoD cells were ordered using genes that were both used to order the HSMM cells (as described in Qiu et al., 2017a, 2017b) and also detectably expressed in hFib-MyoD cells.

(B) Pseudotime kinetics of CDK1, ENO3, and TNNT2 from the root of the trajectory to outcome F_1 (solid line) and the cells up to branch point B_2. Cells are colored by their expression of CDK1 and/or ENO3, as in (A).

(C) All genes expressed in a branch B_1-dependent manner. Each row indicates the standardized (row-centered) kinetic curves of a gene. The center of the heatmap shows the kinetic curve value at the root of the trajectory. Proceeding from the center to the left of the heatmap follows the kinetic curve from the root along the trajectory to outcome F_1. Proceeding to the right yields the curve from the root to B_2. The curve values for each gene are normalized to *Z* scores to facilitate comparison of genes kinetics.

We next analyzed the path that led away from branch point B_1 to a second branch point B_2, which in turn gave rise to outcomes F_2 and F_3 (Figure 4A). The cells at F_2 expressed *MYOG*, lacked markers of active proliferation, and showed upregulation of numerous genes needed for muscle contraction (Figures 4B and S2). Branch-dependent expression analysis at B_2 showed significant changes in 277 genes (FDR < 10%), including a coherent cluster that included *TNNI1*, *TNNT3*, *MYL4*, and other notable proteins involved in myotube contraction (Table S2) (Figure 4C). In contrast, cells at F_3 lacked expression of *MYOG*, cell-cycle genes, and most of those needed for muscle contraction. We therefore concluded that B_1 is a decision point corresponding to whether a cell will exit the cell cycle, while B_2 is a decision point governing whether, having exited the cell cycle, it will go on to activate the myogenic program. Based on our differential analysis, we termed the three hFib-MyoD reprogramming outcomes F_1, F_2, and F_3 as "cell-cycle exit failure," "partial reprogramming," and "reprogramming failure," respectively.

Aligning Single-Cell Trajectories via Dynamic Time Warping

Having identified the core trajectory from the fibroblast state to a partially reprogrammed state, we next sought to quantify the differences between this program of gene expression changes and the one associated with normal myoblast differentiation. In principle, combining the HSMM and hFib-MyoD cells and building a joint trajectory could show how the two cell types converge toward a muscle expression program, with cells of both types co-occupying positions of the graph corresponding to successful activation of the myogenic program. However, Monocle reconstructed a fully linear trajectory with no intermixing of cells at similar stages (Figure S3A). This is likely because despite having some genes regulated in a common direction such as *MEF2C* and *ENO3*, many other genes differ between the cell types, and these differences persist throughout reprogramming and differentiation.

We therefore developed a new algorithmic approach for comparing the HSMM and hFib-MyoD trajectories. A successful comparison between two trajectories should reveal to the user which cells from each are in similar gene expression states and which cells are in states unique to each biological process under study. Because the two single-cell trajectories were learned from each dataset independently, their pseudotime scales are not directly comparable; there is no universal "unit" of pseudotime. Simply combining two different cell types into a single Monocle analysis does not reveal any similarities between their trajectories. We thus devised a "manifold alignment" algorithm (Ham et al., 2003) based on dynamic time warping (DTW) (Vintsyuk, 1972), which matches highly similar points on the two trajectories and creates a mapping between the HSMM and hFib-MyoD pseudotime scales. (Figure 5; Data S1). DTW operates analogously to sequence alignment in that it reports how one trajectory must be deformed (i.e., dilated in some regions and contracted in others) in order to closely match the other trajectory. It matches up the most similar regions of the trajectories so their differences are apparent. In a complementary study published while this manuscript was in revision, Welch et al. devised an algorithm, MATCHER, that can align two linear pseudotemporal trajectories that have been reconstructed from different measurement types (e.g., gene expression and DNA methylation) (Welch et al., 2017). Like MATCHER, our algorithm assumes the cells proceed through linear biological process

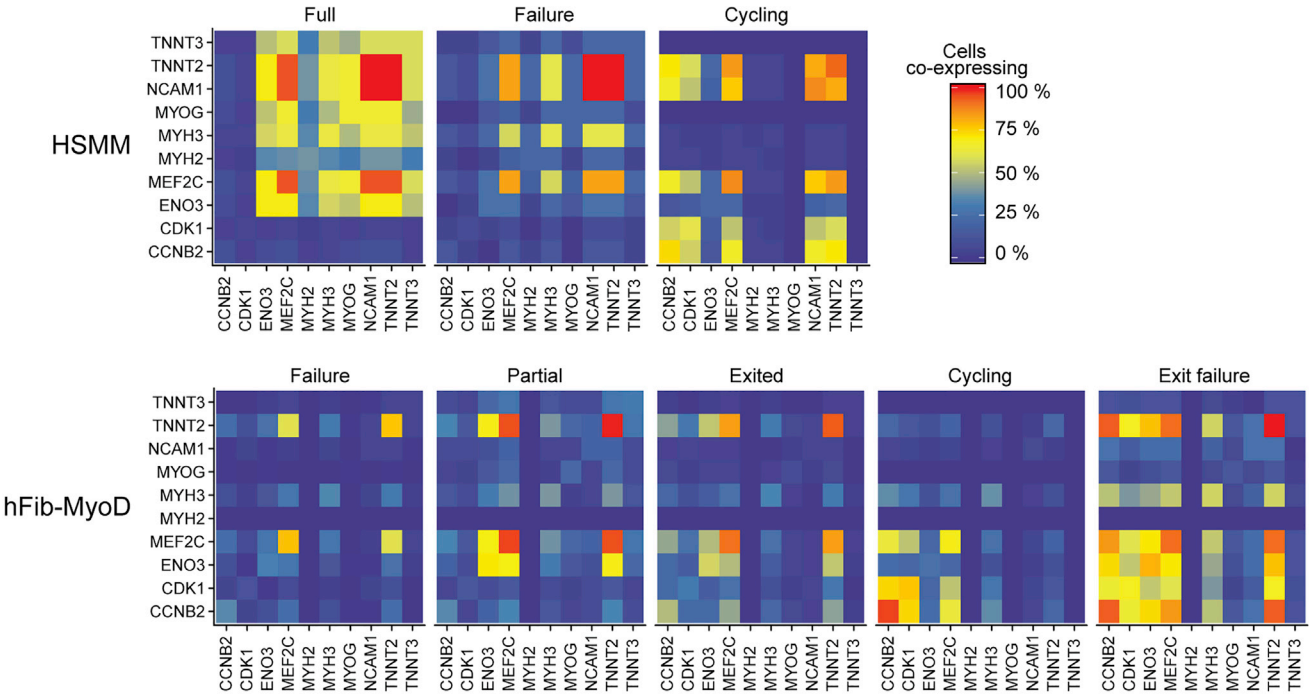


Figure 3. Fraction of Cells Co-expressing Selected Markers of Proliferation and Myoblast Differentiation, Shown for Both HSMM and hFib-MyoD Cells

monotonically and that progression through each biological process leads to measurable changes in gene expression. However, where MATCHER assumes that both experiments examine the same cell type, we aim to align cells of different types (e.g., fibroblasts and myoblasts) that proceed through similar processes.

Our algorithm aligned the HSMM and hFib-MyoD trajectories along a "warping path": a sequence of matchings between an HSMM and an hFib-MyoD cell such that the most transcriptomically similar cells (and thus points in HSMM and hFib-MyoD pseudotimes) are linked (Figure 6A). Just as "edit distance" in DNA sequence alignment measures the number of base transversions, insertions, and deletions needed to transform a sequence into one aligned to it, the length of this warping path quantifies the total amount of transcriptome "distortion" needed to transform one gene expression trajectory into another. Randomly permuting the order of the HSMM and hFib-MyoD cells and then aligning these shuffled trajectories produced far longer warping paths than the correctly ordered trajectories (Figure 6B). Because warping path lengths of shuffled trajectories appeared to be normally distributed ($p = 0.68$, Shapiro-Wilk test), we were able to determine that the warping path between the unshuffled trajectories was significantly shorter ($p < 10^{-16}$, t test), suggesting that DTW fixed on genuine biological signal to match up the pseudotime scales.

Because our method assumes the two input trajectories are correctly inferred, we set out to explore the robustness of trajectory inference with Monocle 2 and its impact on downstream trajectory alignment. Monocle 2 produces trajectories that are robust to many choices of its parameter values, but there are several that can dramatically impact the topology of the trajectory. The set of genes used for ordering, which defines the high-dimensional space in which the original data reside prior to embedding, can have a major impact on the interpretation of the biological process under scrutiny (Qiu et al., 2017a, 2017b). When we used all 441 genes that varied significantly over time in the HSMM experiment instead of just the top 100, Monocle 2 produced a nearly identical HSMM trajectory, but the hFib-MyoD trajectory failed to capture the split between partially reprogrammed and unreprogrammed cells (Figure S3B). Changing one of Monocle 2's key regularization parameters had a similar effect. Ordering both sets of cells by genes that undergo time-dependent changes in expression in hFib-MyoD rather than HSMM led to a loss in the "failure to exit" branch in the hFib-MyoD trajectory and a failure to place some MYOG+ cells on the partially reprogrammed branch (Figure S3C).

The alternative trajectories produced by changing the ordering genes or regularization parameters have some topological features in common with the one shown in Figure 2A, but each failed to segregate the cells of the "reprogramming failure" and "partial reprogramming" outcomes onto distinct paths. Because there is typically no experimentally derived "ground truth" ordering for a set of cells, benchmarking a trajectory's accuracy typically requires checking whether it can be reconciled with time series immunofluorescence or other kinetic measurements of the system. At the 72-hr post-serum switch, both cultures contain nuclei in MYH+ cells as well as cells lacking MYH (Figure 1). Therefore, both trajectories should have at least two outcomes, one with partially reprogrammed cells (MYH+) and one in which the myogenic program failed to activate (MYH−). These analyses suggest that although Monocle 2 is somewhat sensitive to

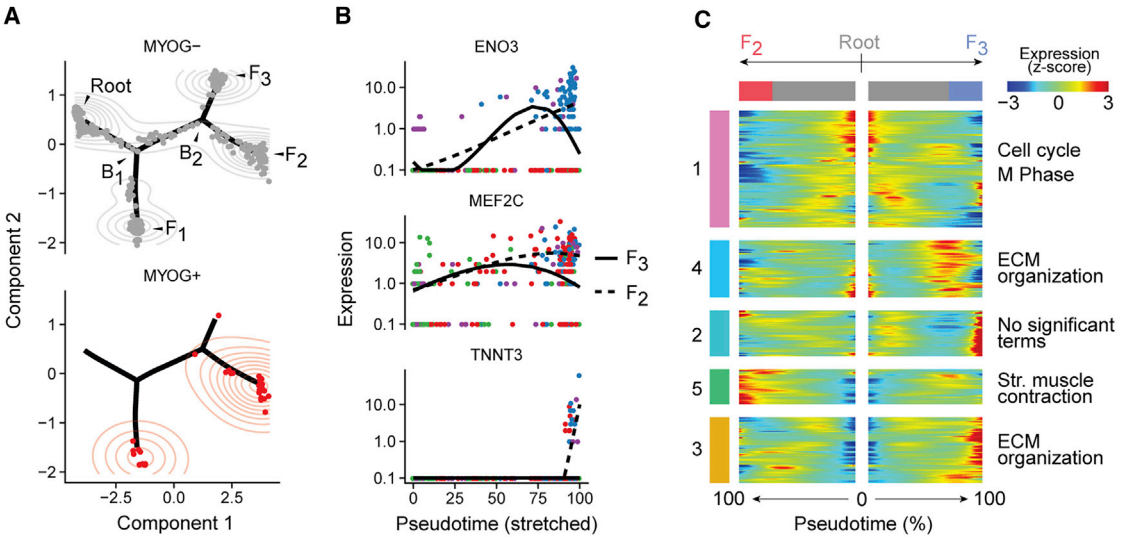


Figure 4. Two of the Outcomes Reached by hFib-MyoD Feature Cell-Cycle Exit Followed by Varying Levels of Activation of the Skeletal Muscle Expression Program

(A) Cells colored by detection of MYOG mRNA.

(B) Kinetic curves for ENO3, MEF2C, and TNNT3 from the root through B_2 to outcomes F_2 (dashed line) and F_3. Cells are colored as in Figure 2A.

(C) Branched kinetic heatmap for all genes dependent on B_2, along with selected enriched gene ontology terms in indicated clusters.

parameters, the trajectories we have aligned here are consistent with the independent techniques for assessing myogenic differentiation kinetics.

Aligning Single-Cell Trajectories Identifies Key Molecular Determinants of Myogenic Reprogramming

Because numerous pathways and transcriptional circuits play a role in muscle development, we therefore sought to use our alignment to reveal which of them might contribute to the failure of hFib-MyoD cells to fully reprogram. The alignment between HSMM and hFib-MyoD revealed extensive quantitative differences in expression kinetics between their trajectories, highlighting several coherent clusters of genes missing in hFib-MyoD or aberrantly regulated relative to HSMM (Figure 6C). Most of the cell-cycle markers, such as *CDK1*, displayed similar kinetics in both hFib-MyoD and HSMM, indicating that on the core hFib-MyoD trajectory, the cell-cycle exit proceeds swiftly following media switch. *MEF2C*, a crucial co-factor for MYOD and MYOG in myoblasts, was upregulated with similar timing but to a lesser extent in hFib-MyoD. *MYOG*, in contrast, was upregulated far later in hFib-MyoD than HSMM, to a lesser extent, and in a smaller proportion of cells (Figure 6D).

In myoblasts, MEF2C and MYOG both auto-regulate their own expression (Edmondson et al., 1992; Wang et al., 2001), after initially being upregulated by MYOD. MYOD activates its targets by forming a heterodimer with E47, which together recruit p300 to robustly transactivate the late myogenic expression program. The activity of p300 at MYOD-bound promoters is driven through Akt-mediated insulin signaling (Serra et al., 2007). Formation of the MYOD/E47 heterodimer is inhibited by ID family proteins (Neuhold and Wold, 1993), which sequester E proteins away from chromatin. In proliferating myoblasts, ID family proteins are maintained at high levels by an incompletely understood mechanism that is likely downstream of BMP signaling (Lewis and Prywes, 2013), which acts to allow sufficient myoblast expansion from satellite cells during regeneration (Sartori et al., 2013). During differentiation, ID expression drops, enabling E47 to form heterodimers with MYOD at promoters of key muscle genes. Expression levels of insulin receptor (INSR) and the insulin-like growth factor receptors (IGF1R and IGF2R) were significantly lower in hFib-MyoD. (Figure 6D) IGF2, which drives myoblast differentiation in an autocrine loop (Florini et al., 1991), was strongly upregulated in HSMMs but not hFib-MyoD. Furthermore, ID1 and ID3 were at far higher levels in hFib-MyoD and were maintained throughout differentiation. BMP4 was expressed in hFib-MyoD, while HSMMs did not express BMP family mRNAs at appreciable levels. (Figure 6D) We thus hypothesized that high levels of BMP and insufficient insulin signaling were locking hFib-MyoD cells in a negative feedback loop, preventing their efficient activation of the myotube expression program.

Cell-Extrinsic Factors Are Required for MYOD-Mediated Myogenic Reprogramming

To test whether modulating insulin or BMP signaling could enhance myogenic conversion efficiency, we supplemented the differentiation medium with recombinant insulin protein, a chemical inhibitor of the BMP receptor, or both. After 120 hr of conversion in the presence of insulin, hFib-MyoD showed marked increases in the number of myotubes (Figure 7A). Myoblasts cultured with insulin did not form significantly more myotubes, but myotube size, the number of nuclei residing in myotubes, and nuclei per myotube all increased at least 2-fold, indicating far more efficient fusion. Inhibition of BMP signaling also significantly increased the number, size, and nuclei count of hFib-MyoD-derived myotubes. (Figure 7B) The addition of both insulin and the BMP inhibitor increased myotube number, size, and nucleation more than adding either alone, suggesting

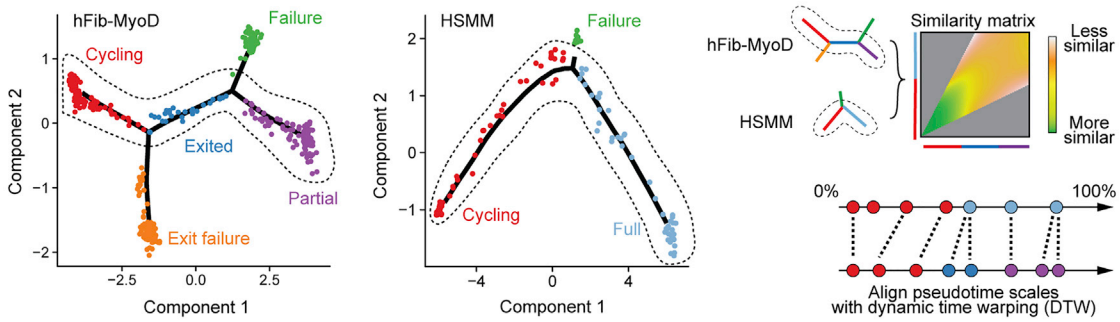


Figure 5. Core Trajectories of Myogenic Reprogramming and Differentiation

The core trajectories of hFib-MyoD (left) and HSMM (center) were aligned using dynamic time warping (Giorgino, 2009; Rabiner and Juang, 1993; Vintsyuk, 1972) (DTW; right), a dynamic-programming time series alignment technique, which matches pseudotime coordinates on the hFib-MyoD trajectory to the most best matching points in the HSMM trajectory as determined by similarity in global expression.

that these pathways act independently to modulate hFib-MyoD conversion efficiency. Importantly, neither insulin nor the BMP inhibitor altered the total number of nuclei in the experiment, indicating that variation in myotube efficiency was not due to variation in cell counts. Together, these results suggest that MYOD is sufficient to convert human foreskin fibroblasts to myotubes only in the presence of appropriate upstream signaling from core myogenic pathways.

MYOD has been characterized as a ''pioneer'' transcription factor because of its sufficiency in remodeling chromatin at inactive regulatory sites in the genome (Zaret and Carroll, 2011). Recently, a number of studies have shown that genetically or chemically modulating the activity of chromatin remodeling enzymes alters the efficiency, timing, and heterogeneity of reprogramming to the pluripotent state. For example, we showed that inhibiting the activity of the histone demethylase LSD1, which removes mono- and di-methyl groups from lysine 4 in the tail of histone H3 (H3K4m1/2), improves fibroblast-to-iPSC conversion efficiency by 10-fold (Cacchiarelli et al., 2015). Similarly, incubating hFib-MyoD cells with an LSD1 inhibitor dramatically increased their conversion to myotubes in the presence of insulin (Figure 6D). In contrast, the inhibitor had only a modest effect in the absence of insulin and no effect on HSMM differentiation.

DISCUSSION

The landmark discovery by Takahashi and Yamanaka that developmental decisions can be reversed with the ectopic expression of four factors, ultimately converting fibroblasts into pluripotent cells, raised the prospect of large-scale manufacture of arbitrary cell types and tissues. (Takahashi and Yamanaka, 2006). During reprogramming, several key developmental decisions are observed (Cacchiarelli et al., 2015; Takahashi et al., 2014). However, for most cell types, we lack effective protocols for differentiating them from pluripotent cells or direct ''reprogramming cocktails'' that would generate them from another cell type. The cocktails that do exist generally only work on one or a handful of initial cell types, often with poor and variable efficiency (Vierbuchen and Wernig, 2011) (Xu et al., 2015). Developing better methods for dissecting the molecular basis for different reprogramming cocktails is therefore critical for realizing the promise of efficient manufacture of therapeutically relevant cells.

MYOD is perhaps the best characterized ectopic reprogramming factor and has been extensively studied as a model ''master regulator'' of cell fate. Central to its identity is its ability to single-handedly convert various cell types into muscle. However, as noted, some cell types, such as HeLa cells (Weintraub et al., 1989), remain refractory to myogenic conversion for reasons that until recently remained poorly understood. Work by Forcales et al. showed that HeLa cells can be converted to myotubes with MYOD only when BAF60C, a subunit of the BAF ATP-dependent chromatin remodeling complex (present in mesodermal lineages), is also expressed along with appropriate signaling from p38α (Forcales et al., 2012). That is, MYOD's ability to reprogram to muscle is dependent upon the presence of factors needed for ATP-dependent chromatin remodeling.

Our study shows that the extracellular signaling environment must also be supportive for efficient MYOD-mediated reprogramming of fibroblasts to myotubes. Using pseudotemporal single-cell transcriptome analysis to ''debug'' our reprogramming system, we identified two molecular barriers to conversion. The first is the failure to engage a central autocrine positive feedback loop driven by insulin signaling, which is supplied by IGF2 in myoblasts. The second is that hFib-MyoD cells secrete BMP4, which may activate a negative autocrine loop and impede myogenic conversion. Supplementing the cellular milieu with insulin and a BMP inhibitor rescued MYOD's ability to convert these fibroblasts to myotubes. Although these pathways are known to play a role in muscle development and reprogramming, so do many others (Knight and Kothary, 2011); trajectory alignment identified these as the specific barriers in our system that should be disrupted to ease conversion. Alignment of single-cell trajectories is thus a powerful technique for isolating the sequence of productive state transitions leading to effective cell type conversion.

Dependence on a supportive extracellular signaling environment might explain variable reprogramming efficiency during not only myogenic reprogramming but many other reprogramming settings. Furthermore, while the importance of engaging regulatory feedback loops for reprogramming, differentiation, and maintenance of the pluripotent and other cellular states is

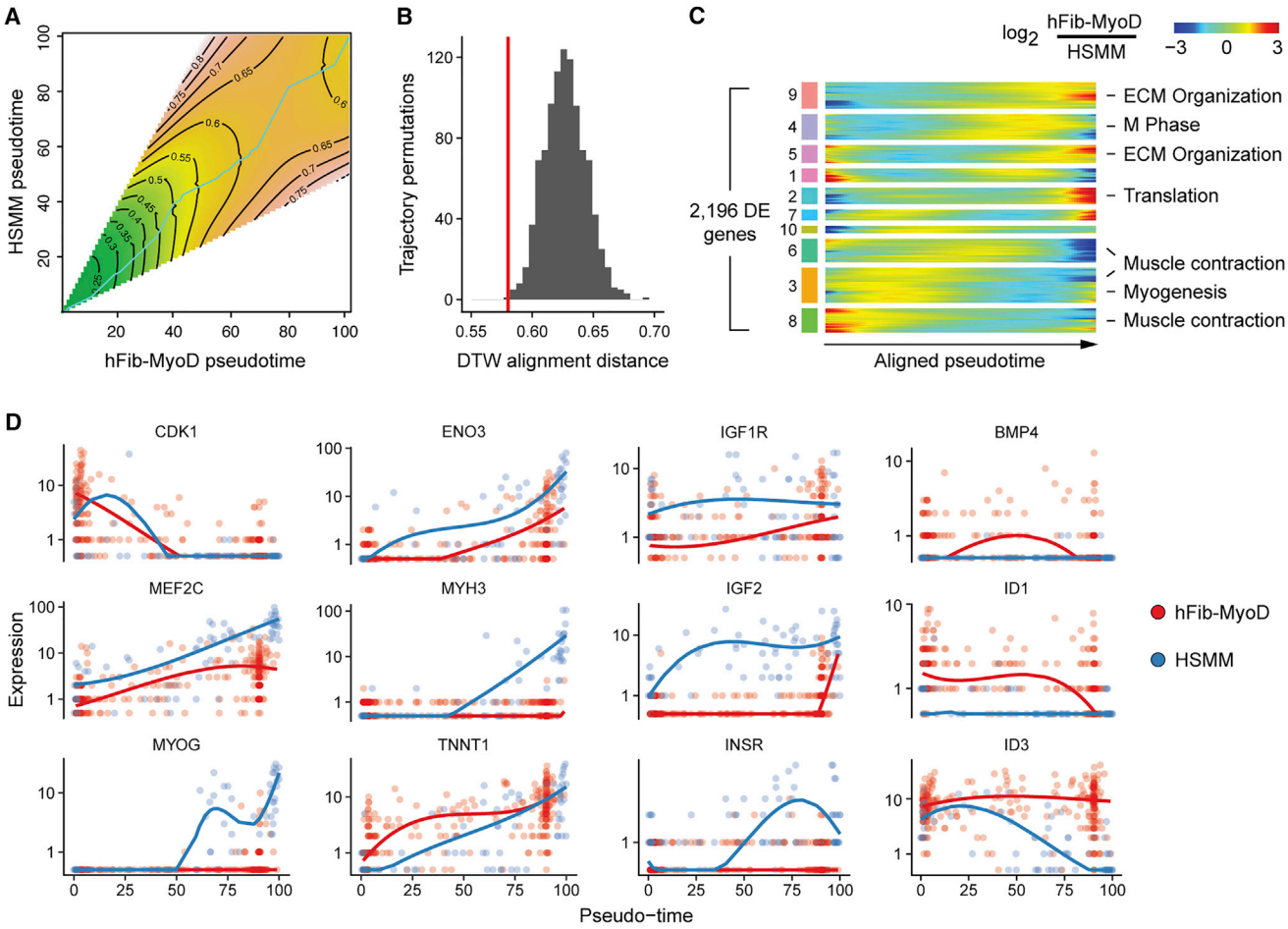


Figure 6. Alignment of Myogenic Reprogramming and Myoblast Differentiation Trajectories Identifies Insulin and BMP Signaling as Aberrantly Regulated in hFib-MyoD

(A) The DTW alignment (cyan curve) follows a dissimilarity-minimizing path through the "landscape" of possible ways to pair up points on the two pseudotime curves. The contours indicate levels of equal similarity between global expression profiles of typical cells at indicated HSMM and hFib-MyoD pseudotimes. HSMM and hFib-MyoD cells were aligned to minimize differences in the union of genes used to order the two sets of cells.

(B) The distribution of normalized lengths of the warping path as reported by the DTW package for shuffled trajectories. Shuffling was conducted immediately prior to computing all pairwise distances between points on the two pseudotime series. The red line indicates the length of the warping path (0.58) of the unshuffled data.

(C) Clustered heatmap of genes showing significant differences in aligned kinetic curves (FDR < 10%; likelihood ratio test; see STAR Methods) after controlling for common pseudotime-dependent differences. Each row shows a different gene. Columns correspond to the log-transformed fold change of the curve for hFib-MyoD over that of HSMM at each point in the aligned trajectory. These values were further transformed into per-gene Z scores prior to clustering. Genes are clustered by Ward's method and tested for enrichment of genes in REACTOME pathways, with selected significantly enriched pathways shown.

(D) Aligned kinetic curves for the indicated marker genes, all of which show significant differential pseudotime-dependent expression.

well appreciated, feedback loops are most frequently discussed in the context of transcriptional circuits that govern these processes. Our work underscores the importance of establishing (or interrupting) autocrine signaling feedback loops in order to reach a desired cellular state.

Beyond providing insights into barriers to myogenic reprogramming, our study shows that pseudotemporal single-cell trajectory alignment is a powerful approach for dissecting biological processes. Alignment can reveal kinetic differences in how each gene is regulated in the two processes and could be useful for understanding not only how reprogramming compares to normal cell differentiation but also how a dynamic process proceeds differently under different environmental conditions or how a mutation alters it. Our alignment algorithm as well as other recently proposed approaches such as MATCHER make several assumptions regarding the underlying biological processes being compared. If these assumptions are not met, or if the trajectories provided as input are incorrectly constructed, the alignment will not yield an accurate picture of how the biological process under study is similar or different. It is conceivable that a future algorithm that jointly learns trajectories for two or more processes simultaneously along with the mapping

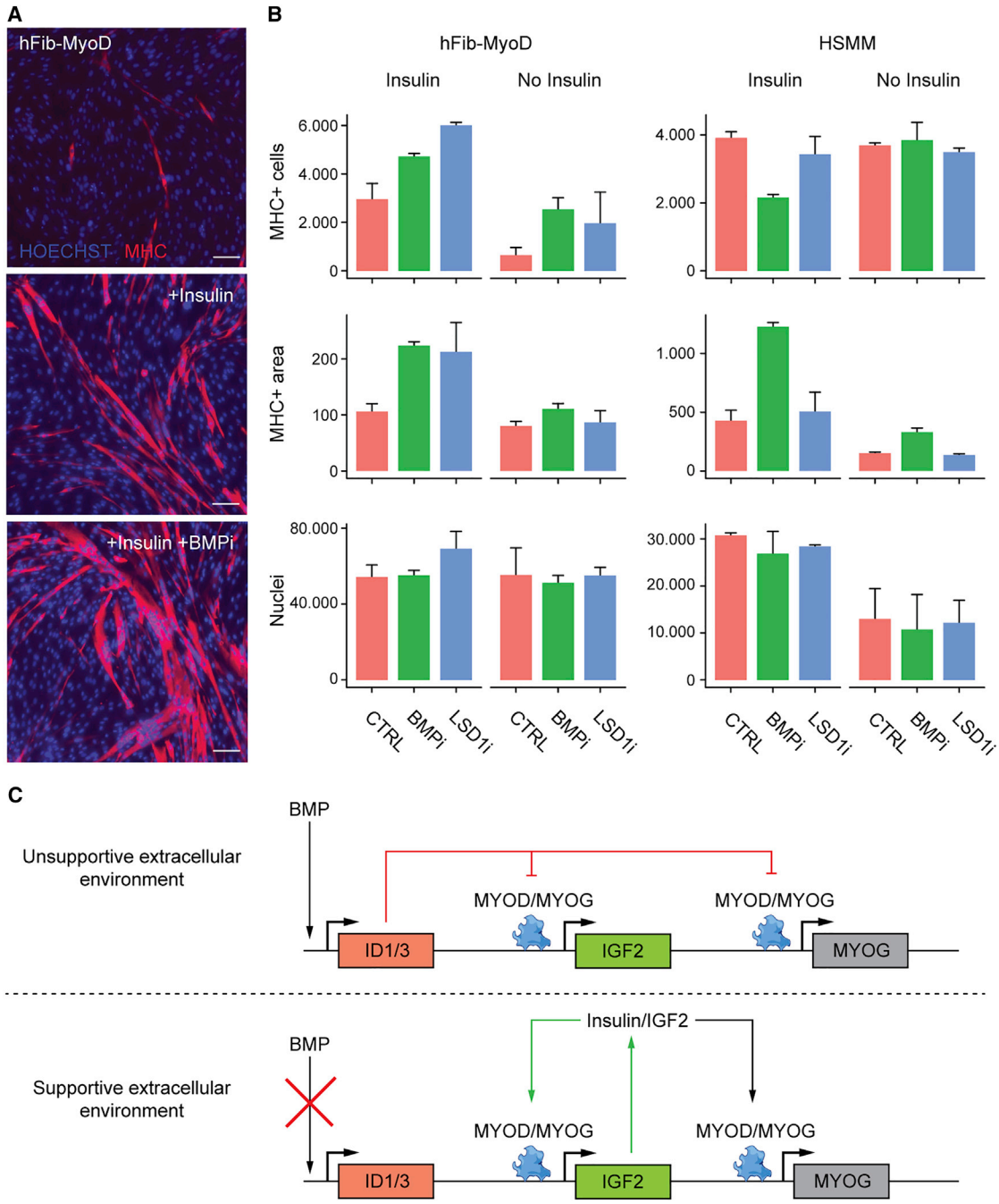


Figure 7. Adding Insulin, Inhibiting BMP Signaling, and Inhibiting LSD1 Activity Increase Myogenic Conversion Efficiency

(A) Immunostaining of hFib-MyoD with muscle-specific anti-myosin-heavy chain (MHC) antibodies 120 hr post induction of MYOD-mediated reprogramming. Top inset indicates control cells. Cells in the middle and bottom insets were treated with insulin or insulin and BMP inhibitor, respectively. Scale bars represent 100 μm.

(B) Counts of MHC+ cells, size of MHC+ cells in pixels, and total nuclei as measured by automated image processing scripts (see STAR Methods). Error bars indicate SD across replicates.

(C) A model summarizing how the extracellular signaling environment conditions cells for reprogrammability by MYOD. An unsupportive environment that exposes cells to BMP maintains cellular ID family protein levels, which in turn suppress the myogenic program by interfering with the formation of active MYOD-containing transcription complexes. A supportive environment lacking BMP and containing insulin engages an IGF2-mediated autocrine feedback loop that amplifies expression of *MYOG* and potentially its co-factors, leading to efficient induction of the myogenic expression program.

between them might be more robust to user-provided parameters. Despite these limitations, trajectory alignment can advance our understanding of not only how two biological processes relate to one another but also how different layers of gene regulation are linked in the same process. Trajectory alignment is applicable to these and many other settings and should yield

insights into the molecular mechanisms of reprogramming, cell differentiation, and a wide array of other biological processes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Culture and Derivation
- METHOD DETAILS
  - Imaging-Based Quantification of Myotube Formation
- SINGLE-CELL TRANSCRIPTOME SEQUENCING
  - Single-Cell Trajectory Reconstruction
  - Trajectory Alignment
  - Differential Expression Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, two tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.cels.2018.07.006.

ACKNOWLEDGMENTS

This work was supported by NIH grants DP2HD088158 to C.T., 1DP2OD00667 (to J.L.R.), and P01GM099117 (to A. Meissner, T.S.M., and J.L.R.) and by Fondazione Telethon Core Grant, Armenise-Harvard Foundation Career Development Award, European Research Council (grant agreement 759154, CellKarma), and the Rita-Levi Montalcini program from MIUR (to D.C.). C.T. is partly supported by an Alfred P. Sloan Foundation Research Fellowship.

AUTHOR CONTRIBUTIONS

C.T. and D.C. conceived the project. X.Q. and C.T. wrote Monocle. D.C., C.T., S.S., A. Manfredi, A.G., J.G., P.P., K.J.L., S.L., and E.O. performed the experiments. C.T. and D.C. led data analysis, with specific analyses performed by M.Z. and D.C., and with assistance from X.Q. C.T. and D.C. wrote the manuscript with assistance from A. Meissner, T.S.M., and J.L.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 24, 2017
Revised: April 3, 2018
Accepted: July 23, 2018
Published: September 5, 2018

REFERENCES

Bentzinger, C.F., Wang, Y.X., and Rudnicki, M.A. (2012). Building muscle: molecular regulation of myogenesis. Cold Spring Harb. Perspect. Biol. *4*, https://doi.org/10.1101/cshperspect.a008342.

Blau, H.M., Chiu, C.P., and Webster, C. (1983). Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell *32*, 1171–1180.

Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., et al. (2015). Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell *162*, 412–424.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Edmondson, D.G., Cheng, T.C., Cserjesi, P., Chakraborty, T., and Olson, E.N. (1992). Analysis of the myogenin promoter reveals an indirect pathway for positive autoregulation mediated by the muscle-specific enhancer factor MEF-2. Mol. Cell. Biol. *12*, 3665–3677.

Florini, J.R., Magri, K.A., Ewton, D.Z., James, P.L., Grindstaff, K., and Rotwein, P.S. (1991). "Spontaneous" differentiation of skeletal myoblasts is dependent upon autocrine secretion of insulin-like growth factor-II. J. Biol. Chem. *266*, 15917–15923.

Forcales, S.V., Albini, S., Giordani, L., Malecova, B., Cignolo, L., Chernov, A., Coutinho, P., Saccone, V., Consalvi, S., Williams, R., et al. (2012). Signal-dependent incorporation of MyoD-BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex. EMBO J. *31*, 301–316.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: the dtw package. J. Stat. Softw. *31*, 1–24.

Ham, J.H., Lee, D.D., and Saul, L.K. (2003). Learning high dimensional correspondences from low dimensional manifolds. In Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining at the 20th International Conference on Machine Learning (International Machine Learning Society), pp. 34–41.

Knight, J.D., and Kothary, R. (2011). The myogenic kinome: protein kinases critical to mammalian skeletal myogenesis. Skelet. Muscle *1*, 29.

Lewis, T.C., and Prywes, R. (2013). Serum regulation of Id1 expression by a BMP pathway and BMP responsive element. Biochim. Biophys. Acta *1829*, 1147–1159.

Neuhold, L.A., and Wold, B. (1993). HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. Cell *74*, 1033–1042.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017a). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017b). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods *14*, 309–315.

Rabiner, L., and Juang, B.H. (1993). Fundamentals of Speech Recognition (PTR Prentice Hall).

Salvatori, G., Lattanzi, L., Coletta, M., Aguanno, S., Vivarelli, E., Kelly, R., Ferrari, G., Harris, A.J., Mavilio, F., and Molinaro, M. (1995). Myogenic conversion of mammalian fibroblasts induced by differentiating muscle cells. J. Cell Sci. *108*, 2733–2739.

Sartori, R., Schirwis, E., Blaauw, B., Bortolanza, S., Zhao, J., Enzo, E., Stantzou, A., Mouisel, E., Toniolo, L., Ferry, A., et al. (2013). BMP signaling controls muscle mass. Nat. Genet. *45*, 1309–1318.

Serra, C., Palacios, D., Mozzetta, C., Forcales, S.V., Morantte, I., Ripani, M., Jones, D.R., Du, K., Jhala, U.S., Simone, C., et al. (2007). Functional interdependence at the chromatin level between the MKK6/p38 and IGF1/PI3K/AKT pathways during muscle differentiation. Mol. Cell *28*, 200–213.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat. Commun. *5*, 3678.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., et al. (2016). Dissecting direct

reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395.

Vierbuchen, T., and Wernig, M. (2011). Direct lineage conversions: unnatural but useful? Nat. Biotechnol. *29*, 892–907.

Vintsyuk, T.K. (1972). Speech discrimination by dynamic programming. Cybern. Syst. Anal. *4*, 52–57.

Wang, D.Z., Valdez, M.R., McAnally, J., Richardson, J., and Olson, E.N. (2001). The Mef2c gene is a direct transcriptional target of myogenic bHLH and MEF2 proteins during skeletal muscle development. Development *128*, 4623–4633.

Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc. Natl. Acad. Sci. USA *86*, 5434–5438.

Welch, J.D., Hartemink, A.J., and Prins, J.F. (2017). MATCHER: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics. Genome Biol. *18*, 138.

Xu, J., Du, Y., and Deng, H. (2015). Direct lineage reprogramming: strategies, mechanisms, and applications. Cell Stem Cell *16*, 119–134.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
MHC/MYH2	eBioscience	Cat. 14-6503; CLONE MF20
Chemicals, Peptides, and Recombinant Proteins		
Puromycin Dihydrochloride	ThermoFisher Scientific	Cat. A1113803
Hygromycin B	ThermoFisher Scientific	Cat. 10687010
Blasticidin S HCl	ThermoFisher Scientific	Cat. A1113903
FGF-Basic	ThermoFisher Scientific	Cat. PHG0263
Doxycycline Hyclate	Sigma-Aldrich	Cat. D9891
Insulin	ThermoFisher Scientific	Cat. 12585014
LSD1 inhibitor - RN-1	EMD Millipore	Cat. 489479
BMP inhibitor - LDN-193189	Stemgent	Cat. 04-0074
Critical Commercial Assays		
C1 Single-Cell Auto Prep System	Fluidigm	Cat. 100-5950
Deposited Data		
hFib-MyoD Reprogramming Single Cell Transcriptome	NCBI Geo DataSets	GEO: GSE105211
Experimental Models: Cell Lines		
Human Foreskin Fibroblasts (BJ)	Stemgent	Cat. 08-0027
Human Skeletal Muscle Myoblasts (HSMM)	Lonza	Cat. CC-2580
Recombinant DNA		
pLVX-EF1α-hTERT-IRES-Puro	This paper	N/A
ViraPower T-Rex Gateway Expression System (MyoD)	This paper	N/A
Software and Algorithms		
Monocle v.2.5.4	N/A	N/A
TopHat v.2.1.1	N/A	N/A
Cufflinks v.2.2.0	N/A	N/A
Human hg19 genome reference	N/A	N/A
Gencode v17 transcriptome index	N/A	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contacts, Cole Trapnell (coletrap@uw.edu) and Davide Cacchiarelli (d.cacchiarelli@tigem.it).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture and Derivation

HSMM derivation, expansion and differentiation was as previously described (Trapnell et al., 2014). Foreskin Human Fibroblasts obtained from commercial vendor (Stemgent) were expanded in alpha-mem supplemented with glutamax, 10% FBS and 16ng/ul of FGF-b (ThermoFisher Scientific). The fibroblasts were then infected with a mixture of lentiviruses encoding hTERT (and Puromycin resistance gene) (pLVX system from Clontech), Tetracycline Repressor (and Geneticin resistance gene) and TR-controlled hMyoD (and Blasticidin resistance gene) (ViraPower T-Rex system from ThermoFisher Scientific). All the resistance genes are constitutively expressed and therefore triple selection was performed and maintained using 1ug/ml of Puromycin, 500ug/ml of Geneticin and 2ug/ml of Blasticidin to generate the hFib-MyoD line.

To perform reprogramming and differentiation experiments HSMM and hFib-MyoD cells were plated in 24-well formats at a density between 50.000 and 100.000 cells per well. Gelatin was sometime used as a coating agent with no significant difference with respect to uncoated dishes. Differentiation and reprogramming was induced using a differentiation media containing alpha-mem supplemented with glutamax and 2% HS (ThermoFisher Scientific), supplemented with 2ug/ml of doxycycline to enact MYOD expression.

When indicated, Insulin was used at 8ug/ml (ThermoFisher Scientific), BMP inhibitor LDN-193189 was used at 0.1uM (Stemgent), LSD1 inhibitor RN-1 was used at 1uM (EMD Millipore).

METHOD DETAILS

Imaging-Based Quantification of Myotube Formation

At the indicated time points, cells were fixed, stained, imaged and processed as previously described (Trapnell et al., 2014). Briefly, we fixed the cells in 4% PFA, stained for and MHC antibody at a 1:500 dilution (ebioscience) and whole-well imaged using a Celigo Cytometer (Nexcelom). The imaging/counting script is available upon request.

SINGLE-CELL TRANSCRIPTOME SEQUENCING

At the indicated time points, cells were harvested by gentle dissociation using TrypLE (ThermoFisher Scientific) and processed for full-length transcriptome sequencing using the Fluidigm C1 Single Cell or bulk mRNA sequencing as previously described (Trapnell et al., 2014). Cells were loaded onto microfluidic chips optimized for capturing cells in 10-17 micron in diameter. Cells were suspended at a concentration of ~250 cells/μl to maximize the number of single cells captured and minimize the occurrence of doublets. In some instances, cells were reloaded if the majority of capture sites were not occupied on the first attempt. As in Trapnell et al. (2014), captured cells were scored by manual on-chip microscopic inspection to determine if they were singletons and free of other debris. Gene expression profiles for each cell were computed using TopHat (v 2.1.1) and Cufflinks (2.2.0) software packages, which were provided with the hg19 genomic reference and GENCODE v17 transcriptome index.

Single-Cell Trajectory Reconstruction

Single-cell pseudotime trajectories were constructed with Monocle version 2.5.4. Monocle 2 is a dramatically more powerful method than our previous software. Briefly, it uses Reversed Graph Embedding, a machine learning technique that given high dimensional data, constructs a principal graph approximating a manifold in a low-dimensional space. Single cells are projected onto this manifold, which orders them into a trajectory and identifies any branch points corresponding to cell fate decisions. (Qiu et al., 2017a)

Myoblasts were ordered similarly to the procedure we described in Qiu et al. Briefly, expression levels were first converted into relative mRNA counts using Census (Qiu et al., 2017b). Cells with more than 1e6 Census RNAs were filtered as cells that failed conversion. From this set, cells with +/- 2 standard deviations more or less than the mean total Census RNAs were also excluded as either low-complexity libraries or suspected doublets not caught by visual inspection. The HSMM culture contains a mixture of myoblasts and contaminating interstitial fibroblasts, so we informatically classified each cell on the basis of semi-supervised clustering.

We then collected a set of ordering genes that defined myoblast differentiation by testing each gene for differential expression between the four time points in the experiment. Selection of ordering genes is crucial for trajectory inference (Qiu et al., 2017a, 2017b). We depart from our previously published protocol for ordering the HSMM data to enrich the ordering genes for those that are key markers of muscle differentiation by selet the top 100 most significant genes for downstream ordering analysis. These genes included MYOG, MEF2C, and other muscle markers, but excluded a number of genes that are highly expressed but less specifically associated with muscle development (e.g. those involved in ribosome biogenesis). Expression profiles were reduced to 2 dimensions using the DDRTree algorithm (Mao et al) included with Monocle 2, via the reduceDimension(), with ncenter=50 and param.gamma=100. These two parameters control how aggressively Monocle 2 regularizes away small branches in the trajectory. They were manually optimized to produce orderings of HSMM and hFib-MyoD that were most consistent with each other and with other data in the study (e.g. immunofluorescence).

hFib-MyoD cells were filtered for quality and ordered similarly to the myoblasts, though we made no attempts to split the population, as there is no contamination of other cells types in the culture. hFib-MyoD cells were ordered on the basis of the same genes used to order the myoblasts, excluding those genes not detected in at least 5 cells. As there are more hFib-MyoD cells than myoblasts, parameters to reduceDimension(), were adjusted (ncenter=100 and param.gamma=100).

Trajectory Alignment

We use Dynamic Time Warping (DTW) to compute the optimal alignment between the HSMM and hFib-MyoD trajectories. We refer readers to (Giorgino, 2009; Rabiner and Juang, 1993) for a detailed discussion, and summarize the algorithm below. Briefly, DTW, aims to compare two longitudinal data series to one another by locally stretching or compressing them and make them resemble each other while preserving their order. The algorithm takes as input a distance matrix d that captures the dissimilarity between each pair of elements of a vector X and Y. Because DTW takes as input the matrix d, rather than directly examining X and Y, they can be vectors of multivariate observations, categorical values, or some mix of these, provided that the user defines a suitable distance function. DTW computes a "warping curve"

$$\phi(k), \quad k = 1, \ldots, T:$$

$$\phi(k) = \left(\phi_x(k), \phi_y(k)\right),$$

where the warping function ϕ_x maps the time indices of X to that of Y and and ϕ_y maps Y indices to that of X. DTW then can compute the alignment cost of a particular warping curve as

$$d_\phi(X,Y) = \sum_{k=1}^{T} d\left(\phi_x(k), \phi_y(k)\right) m_\phi(k) / M_\phi.$$

The per-step weight $m_\phi(k)$ and the normalization constant $M\phi$ aim to make the accumulated distortions along different alignment paths comparable. These weights are determined according to the user's "step pattern," several of which are made available by the DTW implementation. DTW then finds optimal alignment ϕ (subject to additional constraints to preserve the ordering of the data points) via dynamic programming.

To align the hFib-MyoD and HSMM trajectories, we first selected cells along the "productive" paths from the root of each trajectory to outcomes F_2 and F_1, respectively. Next we generated smoothed expression curves (using Monocle 2's genSmoothCurves() function) for the union of the ordering genes used to create each cell type's trajectory. Genes for which curves could not be generated due to numerical instabilities in either cell type were excluded. These curves were then standardized independently, scaling each to have a mean of zero and a standard deviation of 1. The curves were collected into a pair of matrices, H and M for the myoblasts and the hFib-MyoD cells, respectively, with one row per gene and one column for each of the 100 points on each gene's smoothed curve. The pairwise dissimilarities between each of the 100 pseudo time points in M and the 100 points in F were then computed as $d\ (i,\ j) = 1 - cor\ (H_i,\ M_j)$. These pairwise distances capture the global similarity between myoblasts and hFib-MyoD cells at each pair of points on their trajectories. DTW allows the user to specify the cost of skipping points in one time series or another, analogously to how dynamic programming for sequence alignment imposes different costs for insertions, deletions, and mismatches. Here, we provide DTW with a "step pattern" specified by rabinerJuangStepPattern(3, "d"). To project the two trajectories onto a common pseudotime scale, we use the warp() function of the dtw package.

Differential Expression Analysis

We tested for differential gene expression as a function of pseudotime (ψ) as previously described (Qiu et al., 2017b; Trapnell et al., 2014): We fit the following model:

$$\ln(A_i) = \beta_0 + \beta_{\tilde{\psi}} \tilde{\psi},$$

where A_i is the mean of a negative-binomial valued random variable of the Census-normalized transcript count for gene i, and the tilde above ψ indicates that these predictors are smoothed with natural splines during fitting. This model was compared to the reduced model containing only an intercept coefficient

$$\ln(A_i) = \beta_0$$

by likelihood ratio test. Genes with an Benjamini-Hochberg adjusted p-value of less than 0.05 were determined to be dynamic across pseudotime.

When testing for differences between HSMM and hFib-MyoD, we used a modified model to identify genes that were regulated in a pseudotime-dependent manner differently between the two cell types:

$$\ln(A_i) = \beta_0 + \beta_{\tilde{\psi}_{HSMM}} \tilde{\psi} HSMM + \beta_c C + \beta_{C\tilde{\psi}HSMM} C \tilde{\psi} HSMM,$$

where C is an indicator variable encoding the cell type, and $\tilde{\psi}HSMM$ is the pseudotime scale of HSMM, onto which the hFib-MyoD cells were projected by DTW. This full model was compared to

$$\ln(A_i) = \beta_0 + \beta_{\tilde{\psi}HSMM} \tilde{\psi} HSMM.$$

By likelihood ratio test to identify genes with cell-type dependent differences in expression, while controlling for differences in expression due to pseudotime that might be common to the two.

DATA AND SOFTWARE AVAILABILITY

Data for this study have been deposited with the Gene Expression Omnibus at GEO: GSE105211.