UNIVERSITY OF CALIFORNIA SAN DIEGO

# Elucidation of Mechanisms of *in vitro* Myogenesis of Human Induced Pluripotent Stem Cells with Functional Validation *in vitro* and *in vivo*

A dissertation submitted in partial satisfaction of the

requirements for the of degree Doctor of Philosophy

in

Bioengineering

by

Priya Nayak

Committee in Charge:

Professor Shankar Subramaniam, Chair
Professor Shyni Varghese, Co-Chair
Professor Pedro Cabrales
Professor Prashant Mali
Professor Karl Willert

2018

The Dissertation of Priya Nayak is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2018

DEDICATION

For my family.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge the members of both the Varghese and Subramaniam labs with whom I have worked over the last four years, especially my advisors Drs. Shyni Varghese and Shankar Subramaniam. I would also like to acknowledge the members of the Colas lab whose help was crucial for the validation experiments.

Chapter 1 is a literature review to motivate and introduce the themes of the dissertation. “Regulation of the in vitro pluripotent state” Nayak, Priya. The dissertation author was the only author for this section.

Chapter 2 details our findings in “Elucidation of the mechanisms of *in vitro* myogenesis of human induced pluripotent stem cells.” The dissertation author was the only author for this section.

Chapter 3 details our findings in “Elucidation of the mechanisms of *in vitro* myogenesis of human induced pluripotent stem cells.” The dissertation author was the only author for this section.

Chapter 4 is a modified presentation of material that is being prepared for presentation as “Using human induced pluripotent stem cells to create a functional skeletal muscle-on-a-chip” by Agrawal, A, Nayak, P, Varghese, S. The dissertation author was co-author of this material.

Chapter 5 details our findings in “Elucidation of the mechanisms of *in vitro* myogenesis of human induced pluripotent stem cells.” The dissertation author was the only author for this section.

VITA

| | |
|---|---|
| 2011-Present | MD/PhD Candidate, University of California San Diego |
| 2014-2018 | PhD in Bioengineering, University of California San Diego |
| 2007-2011 | Bachelor of Science in Physics, Caltech |

ABSTRACT OF THE DISSERTATION

Elucidation of Mechanisms of *in vitro* Myogenesis of Human Induced Pluripotent Stem Cells with Functional Validation *in vitro* and *in vivo*

by

Priya Nayak

Doctor of Philosophy in Bioengineering

University of California San Diego, 2018

Professor Shankar Subramaniam, Chair
Professor Shyni Varghese, Co-Chair

Skeletal muscle is the most abundant tissue in the body, comprising up to 40% of the total mass. In addition to its most salient role in locomotion, it also plays a central role in metabolic homeostasis. Therefore, injury-, age- and disease-related compromise of skeletal muscle can be extremely detrimental to overall health and quality of life. As such,

the development of regenerative therapies for skeletal muscle that contribute to *in vivo* repair, as well as *in vitro* strategies that can hasten the development of personalized drug treatment, could be of vital importance. Human induced pluripotent stem cells (hiPSCs) are a powerful tool that can meet both these challenges; they are patient-specific, can give rise to derivatives of all three germ layers, including myogenic progenitors, and are scalable since pluripotent cells readily self-renew in vitro. However, the robust, transgene free myogenic differentiation of hiPSCs remains a hurdle.

In this dissertation, I address these challenges by identifying key time-varying signaling, transcriptional, and epigenetic-related mechanisms that lead to enhanced in vitro myogenesis by comparing the longitudinal transcriptomic profiles of multiple hiPSC lines. Furthermore, I show that targeted genetic perturbation at the outset of differentiation may bias lineage specification to the paraxial mesoderm fate. Finally, through the selective expansion and terminal differentiation of hiPSC-derived myogenic progenitor cell populations, we show that they form functional 3D microtissues in an *in vitro* skeletal muscle-on-a-chip platform, as well as give rise to dystophin positive fibers *in vivo* in a murine model of muscular dystrophy.

# Chapter 1: General Introduction and Motivation

## Introduction to Aims

Skeletal muscle is the most abundant tissue in the body, comprising up to 40% of the total mass. In addition to its most salient role in locomotion, it also plays a central role in metabolic homeostasis. Therefore, injury-, age- and disease-related compromise of skeletal muscle can be extremely detrimental to overall health and quality of life. Volumetric loss injuries such as those sustained in accidents or war can overwhelm regenerative capacity, while wasting due to age-related sarcopenia and muscular dystrophies can all result in significant morbidity and disability. On top of individual hardship, diseases of skeletal muscle contribute large costs to the healthcare system, and this cost is projected to grow with increases in life span and overall increase in the average age of the population.

As such, the development of regenerative therapies that contribute to in vivo repair, as well as in vitro strategies that can hasten the development of personalized drug treatment, could be of vital importance. Human induced pluripotent stem cells (hiPSCs) are a powerful tool that can meet both these challenges; they are patient-specific, can give rise to derivatives of all three germ layers, including myogenic progenitors, and are scalable since pluripotent cells readily self-renew in vitro. These cells can contribute to in vivo cell-based therapies either through direct integration with compromised muscle, or through indirect paracrine effects that can enhance endogenous regeneration. Furthermore, hiPSC-derived myogenic progenitors can form the foundation of in vitro platforms for patient-specific tissue- and disease-modeling, and drug testing. One major roadblock, however,

remains the robust, reliable and transgene-free derivation of myogenic progenitors from hiPSCs. In particular, although the mechanisms of terminal myogenic differentiation are well known, the mechanisms that underlie the specification of pluripotent cells to the myogenic lineage in vitro are unknown. We address this issue in the first two aims of this dissertation, before coming full circle in the third aim to validate the in vitro and in vivo function of hiPSC-derived myogenic progenitors.

In the first two aims of this dissertation, we seek to identify the key players and transcriptional networks that regulate the in vitro myogenesis of hiPSCs. Longitudinal data would allow us to make causal mechanistic inferences about upstream versus downstream lineage specification processes. Towards this end, in aim 1 we generated a longitudinal RNA seq dataset from three healthy hiPSC lines, capturing myogenic induction in nine time points, from day zero (undifferentiated) to day 30 (cell-line dependent expression of several markers of terminal specification). Characterization of the progression of differentiation by immunofluorescent staining revealed a relative acceleration in myogenic specification in one line compared to the other two. However, analysis of broad trends in the data showed that the gene expression profiles of the cell lines diverged relatively early in differentiation, before detectable changes in myogenic commitment by immunofluorescence. Furthermore, we implemented a strategy to cluster genes with respect to their time varying expression across cell lines. This led to the identification of modules of genes with biologically relevant functional enrichment. We used these modules for subsequent transcription factor enrichment analysis, as well as to contextualize

unannotated and understudied transcripts, including transcription factors and long noncoding RNAs.

Since initial clustering analysis in aim 1 indicated that gene expression profiles between the cell lines diverged early in differentiation, in aim 2 we hypothesized that there may be key differences in initial germ layer specification at the outset of differentiation. In fact, we observed several differences in the expression of transcription factors related to endoderm versus mesoderm specification- the cell line with more robust myogenesis had preferential expression of genes related to mesoderm specification. We traced the initial differences forward through time by observing the expression of transcription factors that specify lineage in the downstream derivatives of the respective initial germ layers. To understand the key players of lineage specification, we paid particular attention to the composition of chromatin modifying complexes that regulate the exit from pluripotency and the relative expression of key components of major signaling pathways early in differentiation. Finally, we use siRNA knockdown of transcriptional repressors of beta catenin to perturb the gene regulatory network at the outset of differentiation to bias initial germ layer specification to the mesoderm lineage and (perhaps) lead to enhanced myogenic differentiation.

In the final aim, we come full circle to validate the functionality of hiPSC-derived myogenic progenitors in the in vitro and in vivo contexts. First, we modified an existing protocol for the purification, selective expansion and terminal differentiation of myogenic progenitors. We have characterized early- and late-specified progenitor populations based on their expression of key myogenic regulatory factors, and compared the ability of these

populations to both form functional microtissues using an in vitro skeletal muscle on a chip platform, as well as their ability to contribute to in vivo muscle repair in a murine model of muscular dystrophy.

## Literature Review

### Introduction

In the last decade since their advent, human induced pluripotent stem cells (hiPSCs) have been touted as game changers in the field of regenerative medicine. hiPSCs are derived from terminally differentiated somatic cells- usually skin fibroblasts or peripheral blood derived monocytes- by the ectopic overexpression of key transcription factors that both govern the pluripotency gene regulatory network as well as allow for self-renewal in vitro. Since pluripotent cells can give rise to derivatives of all three germ layers, hiPSCs allow for the creation of theoretically unlimited quantities of patient-specific cells of any tissue type, while simultaneously avoiding ethical issues associated with hESC derivation. As such, hiPSCs could be powerful tools for in vivo transplantation and in vitro patient-specific tissue and disease models for drug testing and embryological studies that would be difficult and unethical to perform in vivo. (Avior, 2016). In fact, hiPSCs and ESCs are already used clinically and pre-clinically in a limited number of contexts. In 2016, Takara Bio launched an hiPSC-derived pancreatic beta cell platform for drug screening and diabetes research. Recently, hiPSC-derived dopaminergic neurons ameliorated Parkinson's disease symptoms in a nonhuman primate for two years (Kikuchi, 2017). Phase 1 clinical trials assessed the safety of hESC-derived retinal pigment epithelial cell transplantation for the treatment of macular degeneration (Schwartz, 2014). More recently, autologous iPSC-

derived retinal pigment epithelium was transplanted into one macular degeneration patient, with no adverse effects (Mandai, 2017).

Despite a few successes, a major hurdle to the realization of the untapped potential of hiPSCs remains the reliable and robust differentiation to more terminal tissue types of choice. Although a number of protocols have been published over the years, there is a lack of consensus due to lab-to-lab and cell-line to cell-line discrepancies in the outcome of differentiation. Numerous studies point to both culture conditions and epigenetic regulation as key factors that determine not only differentiation bias, but that also play a role in the regulation of the pluripotent state itself. To better understand the nature of the in vitro pluripotent state it is helpful to consider the lines of thought and inquiry that led to the invention of hiPSCs. We then consider earlier studies carried out in hESCs and mESCs that shed light on the nature of primed versus naïve pluripotency, and that furthered (the still ongoing) research to establish in vitro culture conditions to maintain pluripotent cells. Finally, we review some of the key molecular players in the regulation of cellular identity including chromatin modifying complexes and signaling pathways.

*Historical Perspective*

The invention of hiPSCs in 2007 by Takahashi and Yamanaka (Takahashi, 2007) is perhaps the beginning of the most recent volume in a lengthy and variegated encyclopedia of the study of regeneration. People have long noted both the pathology and potential utility of either unchecked or controlled regeneration, respectively, of human tissues. Ancient Egyptian scrolls make mention of breast tumors and offer treatments including surgery, while Greek texts describe various malignancies as well as their

prognoses (Cancer Atlas). In parallel, people alluded to the regenerative capacity of several organs, most notably the liver as in the story of Prometheus, although it is not clear whether the ancient Greeks had actually observed this medically (Power, 2008). At the time, the etymology of tumors was ascribed to imbalances of the humours or other causes (Open Collections Program, Harvard); several millennia later, the advent of the microscope made possible the beginning of modern pathology, and scientists came to understand that tumors were in fact derived from the same tissues that they were often embedded in or emanated from. The link between cancer and stem cells, first proposed by Virchow in the mid1800s, named embryonic-like cancer stem cells as the originators of malignant tumors (Raggi, 2016). This conceptual link between stem-ness and self-renewal would later become an important engineering constraint in the selection of transcription factors for hiPSC derivation.

By the early and mid-20$^{th}$ century, significant strides had been made to understand and harness the regenerative potential of various tissue types. In particular, the first hepatectomy and visualization of satellite cells in skeletal muscle helped further understanding of the potential of stem cells to rebuild and rejuvenate compromised tissues (Mauro, 1970). Additionally, in the late 1950s, John Gurdon showed that it was possible to reprogram an adult cell to an embryonic state by transferring the nucleus of an adult cell to an enucleated cytoplasm-containing egg. He pioneered this technique, termed somatic cell nuclear transfer (SCNT) to clone a frog (Gurdon, 2003). In the mid 1990s, this technique was used to clone Dolly the sheep (The Life of Dolly), and was recently used to clone non-human primates in China (Liu, 2018). Importantly, the success of this technique

implied the existence of certain- at that time unknown- factors in the cytoplasm that were sufficient to convert an adult cell to a pluripotent state.

Studies of myoblasts in the 1980s underscored the key role of transcription factors in determining cellular identity. Scientists had observed that fibroblasts exposed to an HDAC inhibitor took on a myoblast-like phenotype and could be differentiated into multinucleated cells that resembled myotubes. They screened several candidates that resulted in the identification of a protein, which they named MyoD for myogenic differentiation factor (Davis, 1987). Overexpression of this protein in fibroblasts was sufficient to convert these cells into myoblasts in the first published demonstration of transdifferentiation of one terminally differentiated cell to another lineage altogether. Taken together, the knowledge that transcription factor overexpression was sufficient to reprogram a cell to another fate, which included a pluripotent state, as well as the significant benefits that could be had with a self-renewing, pluripotent population in vitro set the stage for the invention of hiPSCs.

In 2006 Takahashi and Yamanaka published their derivation of mouse induced pluripotent stem cells, followed by their 2007 publication of the derivation of human induced pluripotent stem cells. A key insight was the inclusion of transcription factors that promoted both pluripotency and self-renewal. For this discovery, Yamanaka split the 2012 Nobel Prize in physiology and medicine with Gurdon, emphatically underscoring the clear thread between SCNT and iPSC derivation. Since 2007, technological advances have given rise to several strategies for hiPSC derivation, including the now widely used Sendai virus (Malik, 2013). However, in addition to relatively low efficiency of reprogramming, there

can be significant heterogeneity of the pluripotent state which could affect their differentiation, including variable expression of key transcription factors such as Nanog (Singh, 2007). While hESC lines are considered the gold standard for in vitro pluripotency, there may be significant heterogeneity in their culture (Drukker, 2012). These variabilities are heightened in hiPSCs; on top of the baseline heterogeneity, hiPSCs may also have persistence of epigenetic memory from before reprogramming (Polanco, 2013). Therefore, a better understanding of the regulation of pluripotency would be helpful.

*Epigenetic control of the pluripotent state*

A number of differences in genome regulation and chromatin structure can account for some of the heterogeneity in pluripotent cell behavior in vitro. Two distinct but interconvertible pluripotent states have been identified, first in mouse ESCs and later in human ESCs. Mouse pluripotent stem cells are grouped into two categories: "ground" state and "primed" pluripotent states (Nichols, 2009). These correspond most closely to the pre-implantation inner cell mass and the post-implantation epiblast, respectively, in terms of their epigenetic state and expression of key transcription factors. The ground pluripotent state exhibits homogeneous expression of key pluripotency transcription factors, including Nanog and Oct4; there is global DNA hypomethylation corresponding to a "naïve" or "blank slate" epigenetic state, with lineage specifying genes firmly turned off. Additionally, in female cells, both X-chromosomes are active implying that the ground state corresponds to an epigenetic state that precedes X-inactivation. By contrast, colonies of hPSCs in the "primed" state have heterogeneous expression of key transcription factors and lineage specification genes are marked with bivalent histone marks, signifying greater

readiness to differentiate. In particular, the bivalent mark consists of simultaneous repressive mark (H3K27me3) plus an activating mark (H3K4me3) (Voigt, 2013). In female cells, only one X-chromosome is active, and there are differences in the binding and genome regulation of the key transcription factors, most notably Oct4. Although mouse ESCs can be derived readily as either ground or primed states, human ESCs are derived only into the primed state. Some reasons for this could be differences in early embryonic development between mice and humans. Since "ground" state pluripotency is more homogeneous and more naïve than the primed state, it is more desirable for differentiation. Studies with mouse embryonic stem cells (mESCs) laid the groundwork for the establishment of culture conditions for the maintenance of pluripotent cells in vitro. mESCs were initially cultured on a feeder layer of inactivated mouse embryonic fibroblasts (MEFs) with a medium that was supplemented with serum. However, use of animal components and cells are suboptimal for clinical translation due to potential immunogenicity and batch-to-batch variability. MEFs were found to secrete leukemia inhibitory factor (LIF) that promoted self-renewal in culture by blocking STAT3. In addition, the pluripotent state could be maintained with media containing LIF supplemented with a cocktail of chemical inhibitors of several signaling pathways; serum could be replaced with knock-out serum replacement. The most widely used inhibitors are CHIR99021 (GSK3B inhibition) and PD0325901 (MEK inhibition) (2i+LIF), or 2i with an additional FGF inhibitor added (3i+LIF). This composition could both maintain as well as select for ground state mESCs in culture. (Tamm, 2013) By contrast, primed mESCs as well as hiPSCs, were shown to have reduced survival in 2i+LIF medium; their survival and self-renewal instead required

media supplemented with FGF and activin. These differential culture conditions reflect the transcriptomic and epigenetic differences between the ground and primed pluripotent states.

Standard maintenance and expansion culture conditions for hPSCs call for a feeder layer or ECM-coating (previously inactivated MEFs, but now more often Matrigel or a fully defined Vitronectin-based ECM) (Badenes, 2016) with KOSR-supplemented medium containing N2 and bFGF supplements. More recently, feeder free culture conditions have become widely prevalent using Matrigel and mTeSR maintenance medium (Ludwig, 2007). However, these standard, feeder-free conditions maintain hPSCs in a state closer to the mouse primed state rather than the ground state.

As the primed state could lead to biased differentiation, several studies have sought to establish culture conditions that are able to convert hESCs and hiPSCs to the ground state by ectopic expression of key transcription factors or targeted manipulation of key signaling pathways and epigenetic regulators. Takashima et al (Cell 2014) show that over-expression of NANOG and KLF2 in hPSCs gives rise to global hypomethylation and a transcriptome that more closely aligns with the mouse ground pluripotent state. Theunissen et al developed a 5i medium condition, analogous with the 2i condition in mouse pluripotent cells, which also allowed for the resetting of primed pluripotency to a state resembling the ground state: loss of bivalent epigenetic marks at lineage specifying genes, increased use of Oct4 distal enhancers, and a shift in the transcriptome. In 2017, Guo et al showed that treatment with a histone deacetylase inhibitor (HDACi) is able to reset hPSCs to a ground-like state. Currently, a commercially available kit (NaiveCult, STEMCELL

Technologies), developed in cooperation with the authors of Guo et al, uses a HDACi and a series of defined media to reset hPSCs to a naive state. Compared to standard hPSC culture and propagation, this is more expensive and more involved as it requires an irradiated MEF feeder layer and careful passaging; however, it may herald the beginning of more standardized hPSC populations for future studies. More study is required that compares the differentiation biases - or lack thereof- of reset ground state hPSCs to that of standard "primed" hPSCs. Altogether, these studies that focus on the conversion between pluripotent states highlight the dynamic nature of an intricate network that integrates multiple signals such as signaling pathways, transcriptomic changes, and multiple forms of epigenetic regulation to achieve pluripotent states.

*Chromatin modifying complexes and pluripotency*

It is also useful to look into various executors of epigenetic regulation that play a role in the regulation of the pluripotent state. Epigenetic modifications at the histone level, such as post-translation modification of the histone protein components or modification of histone tails, as well as modifications at the nucleosome level involving ATP-dependent processes that remove and add histones to chromatin have been shown to play a role in development and the regulation of pluripotency. Modifications at the histone level include acetylation/deacetylation and methylation/demethylation (Zhang, 2016).

Polycomb Repressor Complexes

The Polycomb Repressor Complexes (PRC1/2) are multiprotein complexes with variable composition that play crucial roles during embryogenesis and development, as evidenced by murine PRC2-knockout embryonic lethality. The Polycomb complexes were

first studied in Drosophila; mutations in the Polycomb group proteins led to an abnormal segmented body plan, from whence the name Polycomb came. (Margueron, 2011) The core enzymatic components of PRC1 are ubiquitin ligases RING1 and RNF2 (Ring1B). PRC1 ubiquitinylates histone A2 at lysine 119 (H2AK119ub1), and is generally thought to repress transcription at the sites to which it is targeted. PRC1 has several canonical and noncanonical forms (Conway, 2017) described, with corresponding implication for differential targeting and mechanism of function. In canonical PRC1, the CBX subunits can serve as histone methylation readers. In particular, they target PRC1 to H3K27me3 marks deposited by PRC2 to turn off gene transcription at these sites. However, several studies show that noncanonical PRC1 complexes can bind DNA independently of PRC2 methylation; in fact, they may be the founding members of PRC-mediated silencing in these cases. Studies of noncanonical PRC1 in leukemia indicate that PRC2-independent gene silencing of several important metabolism and cell cycle genes by ncPRC1 promote carcinogenesis (van den Boom, 2015).

In addition, transcription factors can also recruit PRC1 to chromatin to mediate gene silencing. Dietrich et al (Dietrich, 2012) show that PRC1 and REST co-localize at the promoters of genes that regulate neuronal development, and are displaced during neuronal maturation. In pluripotent cells, the lysine demethylase KDM2B is recruited to CpG islands of developmental genes (He, 2015), and in turn recruits a noncanonical PRC1 variant to mediate gene silencing. KDM2B-mediated ncPRC1 recruitment can also regulate reprogramming to the pluripotent by the silencing of key lineage-specifying genes, including SOX17. (Zhou..Chen, Cell Reports 2017) The changeable components of PRC1

have also been directly linked to the regulation of pluripotency: PCGF6 overexpression has been shown to maintain pluripotency in ESCs despite LIF withdrawal (Yang..Rana SciRep, 2015); however, many genomic sites occupied by PCGF6 were not H2AK119ub1or bound by RNF2, so perhaps in this case PCGF6 is acting independently of PRC1. However, several studies have noted a switch in the pattern of CBX component expression from pluripotent ESCs (CBX7) to differentiating ESCs (CBX4, CBX2) (Morey et al 2012, O'loghlen et al 2012, Alogia, 2013)

PRC2 has four core components: H3K27 methyl transferases EZH1/2 that are mutually exclusive in the fully formed complex, SUZ12, EED, and either RBAP46 or RBAP48. EZH1/2 autoinhibit methyltransferase activity when uncomplexed, so PRC2 must be a fully formed complex to have methyltransferase activity. Interestingly, one study (Chamberlain, Yee, Magnuson) reports that PRC2 function is dispensable in the maintenance of pluripotency. EED null ES cells are viable and participate in chimeric embryos, demonstrating their pluripotency; however, embryos containing a high proportion of EED null cells are not viable beyond 10.5 days. This indicates that although adequate expression of core pluripotency factors is sufficient for the maintenance of pluripotency, PRC2 plays an important role in mediating lineage specification upon induction of differentiation

In addition to the four core PRC2 components, several ancillary proteins have been shown to associate with and regulate the function and/or targeting of PRC2 in various contexts. JARID2 and MTF2 and described to be highly expressed in mouse embryonic stem cells, and their knock down facilitates differentiation by altering PRC2-mediated

H3K27me3 marks. (Zhang..Wang, 2011 Stem Cells). JARID2 may also effect the regulation of pluripotency by promoting Nanog expression (Landiera, Fisher 2015) through the targeting of PRC2 to unmethylated CpG regions. PRC2 can also be targeted to the genome through Polycomb repressive element (PREs), which are repeating DNA motifs that may be involved in trans gene regulation through the formation of DNA loops (March, 2017). In addition, the C2H2 zinc finger family of transcription factors was highly enriched in a study of PRC2 binding in Arabidopsis (Xiao, 2017).

There are several additional means of targeting for both PRC1 and 2 in addition to complex composition-mediated binding, recruitment to CpG islands, or interaction with transcription factors, most notably PRC interaction with noncoding RNA. One important example of ncRNA-mediated gene silencing through recruitment of both PRC complexes is the process of X inactivation. The ncRNA Xist initiates X inactivation by covering the DNA and recruiting chromatin remodeling complexes. Although both PRC1 and 2 are normally recruited, and can recruit each other, a study of PRC2 knockdown shows that Xist was able to recruit PRC1 via RNF2 instead of through Eed (Scheftner, Wutz 2006). Furthermore, Xist expression was able to establish epigenetic memory, and allowed for the efficient H2AK119ub1 via PRC1 in downstream cells. Another example is recruitment of PRC2 to the HOX gene family by the ncRNA HOTAIR. Deletion of HOTAIR leads to derepression of key regulator genes and leads to homeotic transformation in the developing embryo (Li, Chang, Cell Reports 2013). Interestingly, although ncRNAs can lead to PRC2 recruitment and gene silencing, nascent RNAs can also interact with PRC2 and have an opposite effect: The nascent RNA can bind PRC2 and shield the DNA from its methylase

activity, thus promoting gene expression. (Kaneko..Reinberg 2014; Blackledge, Rose, Klose Net Rev MCB 2015)

PRC1 and 2 play an important dual role to maintain the poised, bivalent epigenetic state of key lineage specification genes during primed pluripotency. The bivalent domain refers to overlapping H3K27me3 (repressive modification) and H3K4me3 (activating modification) that have been detected in ESC chromatin. These domains are poised since they would only have to lose the h3K27me3 mark to initiate transcription. Ku et al. (PLOS Genetics, 2008) used high throughput sequencing to characterize two types of bivalent domains: one that is PRC1 positive, and one that is PRC1 negative. The domains that have PRC1 recruitment are observed to be more likely to maintain their epigenetic H3K27me3 modification upon differentiation, while the PRC1 negative bivalent domains were not. They suggest that positive PRC1 recruitment may depend on the absence of other activating transcription factors, and that PRC1 presence promotes retention of epigenetic memory upon cell division. Thus, the dual roles of PRC1 and 2 control the activation of genes that specify lineage, and also play a role in the persistence of epigenetic memory.

ATP-dependent chromatin remodeling complexes

The ATP-ase class of chromatin modifiers (SWI/SNF, ISWI, CHD, INO80) mediate changes in nucleosome level organization of chromatin. As such, they are best known for their roles in DNA replication and repair. However, several of these complexes have been implicated in carcinogenesis and stemness.

The SWI/SNF (switch/sucrose non-fermentable) complexes, originally purified from yeast, are highly-conserved, multiprotein, ATP-dependent chromatin remodeling

complexes. The complex consists of one of two mutually exclusive ATP-ases: SMARCA2 (Brm) or SMARCA4 (Brg1), and 8-11 additional protein subunits; there are two main complexes: BAF and PBAF, which in turn have changeable final compositions depending on their cellular context. (Wilson et al. Nat Rev Cancer 2011). Although the SWI/SNF are considered part of the Trithorax group of chromatin remodelers (Schuettergruber et al.) that tend to oppose the transcriptional repression of the PRCs, SWI/SNF complexes may also repress transcription as well. The positive regulation of transcription by SWI/SNF is due to ejection of nucleosome from chromatin, leaving the DNA more accessible to RNA PolII and transcription factors.(Lorch, Kornberg, PNAS 2010) However, SWI/SNF complexes may also recruit HDACs to chromatin, leading to deacetylation and therefore tighter binding of DNA to the nucleosomes. (Zhang et al. Mol Cell Bio, 2002).

Several studies have linked mutations in BAF or PBAF components to carcinogenesis, highlighting their composition-specific function. PBRM1 (Baf108) subunit mutations have been identified in 41% of renal cell carcinomas (Varela, Nature 2011); ARID1A subunit mutations have likewise been identified in 50% of ovarian clear cell carcinomas and 30% of endometrioid carcinomas (Jones et al Science 2010, Wiegand et al NEJM 2010). A recent study (Guerrero-Martinez, Reyes Sci Rep 2018) has linked high expression of SMARCA2 or SMARCA4 with either good or bad prognosis, respectively, in liver and kidney cancer.

Similarly, an ESC-specific BAF (esBAF) has been identified. (Ho..Crabtree PNAS 2009). esBAF is contains SMARCA4 as its core ATP-ase component; knockdown of this component led to decreased in vitro proliferation of ESCs. Likewise, knockdown of

specific esBAF components led to loss of Oct4 expression and pluripotency. Finally, proteomic analyses revealed that esBAF directly interacted with several key transcriptional regulators of pluripotency, including SOX2 and JARID2. Upon differentiation, BAF and PBAF components that were absent or reduced in pluripotent cells have increased expression. (Kaeser, Emerson, J Biol Chem 2008), and purification of BAF complexes containing these specific subunits were shown to have different biochemical properties than other BAF compositions. Furthermore, Brg1 was shown to occupy the promoters of several key regulators of pluripotency, including Nanog, Oct4, Sox2 as well as several polycomb group proteins. (Kidder, Palmer, Knott, Stem Cells 2009). In addition to a pluripotency-specific composition, the SWI/SNF complexes also have crucial tissue specific compositions: Baf60c is specifically expressed in the developing mouse heart and somites (Lickert, Bruneau, Nature 2004) and a switch in SWI/SNF composition was shown correlate with neuronal differentiation; impairing this switch impaired differentiation (Lessard, Crabtree Neuron 2007).

The NuRD (Nucleosome Remodeling and Histone Deacetylation) complex was initially discovered in 1995 when the core ATP-ase units were found to be the autoantigens in dermatomyositis. (Denslow, Wade Oncogene 2007) As its name suggests, it has both post-translational histone modification capability through its association with HDAC1/2, as well as ATP-dependent nucleosome remodeling capability through the CHD3/4 subunits (Mi-2a/B). (Ramirez, Hagman 2009). NuRD complexes are also known to bind methylated CpG islands and work with other chromatin remodeling complexes including PRC2 to mediate transcription regulation. NuRD complex activity has been linked to the seemingly

contradictory roles of attenuating the expression of both the key genes that regulate pluripotency (Reynolds 2012) as well as coordinating with PRC2 to maintain the repressive H3K27me3 repressive mark at bivalent domains upstream of key lineage specification genes (Whyte 2012). NuRD complexes also oppose esBAF-mediated pro-transcriptional action (Yildrim 2011). As a result, it is suggested that NuRD complexes function to “fine tune” gene expression levels at the exit of pluripotency.(Hu and Wade 2012).

As in the other chromatin modifying and remodeling complexes, NuRD components exhibit a context-specificity. The MBD subunits are mutually exclusive: MBD2 is able to bind methylated DNA, while MBD3 is not. (Lai, Wade 2011). The MTA components show preferential expression in both grade and tissue of origin of a number of cancers. For example, MTA1 is associated with higher-grade tumors in all tissue types, including breast cancers. By contrast, MTA3 expression was associated with maintenance of normal breast tissue, perhaps by repressing the gene Snail, which promotes epithelial to mesenchymal transition (EMT). (Fujita, Wade Cell. 2003). In pluripotent cells, NuRD complexes containing MBD3 play a role in regulating the exit from pluripotency, as evidenced by mESCs deficient in MBD3 that maintained pluripotency in the absence of LIF(Kaji..Hendrich 2006 Nat Cell Bio). Recently, an ESC-specific NuRD complex was identified, in which an isoform of MBD3 that is specifically expressed in ESCs, MBD3C, interacts with WDR5 for targeting to genes that regulate pluripotency. (Ee, Fazzio 2017 Stem Cell Reports)

*Embryological Perspective*

Human pluripotent cells in vitro are thought to be analogous to the primed epiblast of the bilaminar blastocyst. Therefore, an understanding of the in vivo developmental processes by which the primed epiblast gives rise to endodermal and mesodermal progenitors may help understand initial germ layer commitment in vitro. The bilaminar blastocyst consists of the primed epiblast as well as the hypoblast. Paracrine signals from the hypoblast and the extraembryonic tissue guide the priming of the epiblast and its subsequent differentiation. The early primitive streak forms in the posterior epiblast. Cells from the epiblast then give undergo epithelial to mesenchymal transition (EMT) and migrate through the primitive streak, into the hypoblast area and give rise to endoderm progenitors, head mesenchyme, cardiogenic mesoderm, lateral plate mesoderm, intermediate mesoderm, and paraxial mesoderm, as well as neural crest progenitors. The anterior early primitive streak gives rise first to endodermal progenitors. Next, cardiogenic mesoderm progenitors/head mesenchyme migrate through the primitive streak in an anterior manner; portions of the head mesenchyme give rise to both cardiac progenitors as well as the skeletal muscles of facial expression. This mesenchyme is unsegmented, in contrast to the segmented somitic mesoderm, a derivative of the paraxial mesoderm) that gives rise to the muscles of the trunk and limbs. This is an important distinction because diseases of skeletal muscle are shown to have differential effects in head versus limb or trunk skeletal muscle. In addition, the transcriptional circuitry that regulates the initial specification of skeletal muscle progenitors may vary between head versus trunk versus

limb muscles, although terminal differentiation proceeds in a similar fashion, via the canonical pattern of myogenic regulatory factor (MRF) expression.

The late, posterior primitive streak gives rise to the three main mesoderm subtypes: lateral plate mesoderm, which contributes to the formation of the cardiovascular system, intermediate mesoderm, which gives rise to the nephric system, and paraxial mesoderm, which gives rise to skeletal muscle and the dermis. The formation of the primitive streak, followed by the mesoderm, first established the anterior/posterior axis of the developing embryo, and next leads to the trilaminar disc.

**References**

Avior, Y., Sagi, I., Benvenisty, N., 2016. Pluripotent stem cells in disease modelling and drug discovery. Nature Reviews Molecular Cell Biology 17, 170–182.

Badenes, S.M., Fernandes, T.G., Cordeiro, C.S.M., Boucher, S., Kuninger, D., Vemuri, M.C., Diogo, M.M., Cabral, J.M.S., 2016. Defined Essential 8™ Medium and Vitronectin Efficiently Support Scalable Xeno-Free Expansion of Human Induced Pluripotent Stem Cells in Stirred Microcarrier Culture Systems. PLOS ONE 11, e0151264.

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

Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Södersten, E., Hansen, K., 2012. REST–Mediated Recruitment of Polycomb Repressor Complexes in Mammalian Cells. PLOS Genetics 8, e1002494.

Drukker, M., Tang, C., Ardehali, R., Rinkevich, Y., Seita, J., Lee, A.S., Mosley, A.R., Weissman, I.L., Soen, Y., 2012. Isolation of primitive endoderm, mesoderm, vascular endothelial and trophoblast progenitors from human pluripotent stem cells. Nature Biotechnology 30, 531–542.

Guo, G., Meyenn, F. von, Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., Plath, K., Smith, A., 2017. Epigenetic resetting of human pluripotency. Development 144, 2748–2763.

Gurdon, J.B., Byrne, J.A., 2003. The first half-century of nuclear transplantation. PNAS 100, 8048–8052.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., Zhang, Y., 2013. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat. Cell Biol. 15, 373–384.

History of Cancer [WWW Document], n.d. . The Cancer Atlas. URL http://canceratlas.cancer.org/history-cancer/ (accessed 5.22.18).

Kikuchi, T., Morizane, A., Doi, D., Magotani, H., Onoe, H., Hayashi, T., Mizuma, H., Takara, S., Takahashi, R., Inoue, H., Morita, S., Yamamoto, M., Okita, K., Nakagawa, M., Parmar, M., Takahashi, J., 2017. Human iPS cell-derived dopaminergic neurons function in a primate Parkinson's disease model. Nature 548, 592–596.

Liu, Z., Cai, Y., Wang, Y., Nie, Y., Zhang, C., Xu, Y., Zhang, X., Lu, Y., Wang, Z., Poo, M., Sun, Q., 2018. Cloning of Macaque Monkeys by Somatic Cell Nuclear Transfer. Cell 172, 881–887.e7.

Ludwig, T., A Thomson, J., 2007. Defined, feeder-independent medium for human embryonic stem cell culture. Curr Protoc Stem Cell Biol Chapter 1, Unit 1C.2.

Malik, N., Rao, M.S., 2013. A Review of the Methods for Human iPSC Derivation. Methods Mol Biol 997, 23–33.

Mandai, M., Watanabe, A., Kurimoto, Y., Hirami, Y., Morinaga, C., Daimon, T., Fujihara, M., Akimaru, H., Sakai, N., Shibata, Y., Terada, M., Nomiya, Y., Tanishima, S., Nakamura, M., Kamao, H., Sugita, S., Onishi, A., Ito, T., Fujita, K., Kawamata, S., Go, M.J., Shinohara, C., Hata, K., Sawada, M., Yamamoto, M., Ohta, S., Ohara, Y., Yoshida, K., Kuwahara, J., Kitano, Y., Amano, N., Umekage, M., Kitaoka, F., Tanaka, A., Okada, C., Takasu, N., Ogawa, S., Yamanaka, S., Takahashi, M., 2017. Autologous Induced Stem-Cell–Derived Retinal Cells for Macular Degeneration. New England Journal of Medicine 376, 1038–1046.

March, E., Farrona, S., 2017. Polycomb silencing mediated by specific DNA-binding recruiters. Nature Genetics 49, 1416–1417.

Margueron, R., Reinberg, D., 2011. The Polycomb complex PRC2 and its mark in life. Nature 469, 343–349.

Nichols, J., Smith, A., 2009. Naive and Primed Pluripotent States. Cell Stem Cell 4, 487–492.

Open Collections Program: Contagion, Humoral Theory [WWW Document], n.d. URL http://ocp.hul.harvard.edu/contagion/humoraltheory.html (accessed 5.22.18).

Polanco, J.C., Ho, M.S.H., Wang, B., Zhou, Q., Wolvetang, E., Mason, E., Wells, C.A., Kolle, G., Grimmond, S.M., Bertoncello, I., O'Brien, C., Laslett, A.L., 2013. Identification of Unsafe Human Induced Pluripotent Stem Cell Lines Using a Robust Surrogate Assay for Pluripotency. STEM CELLS 31, 1498–1510.

Power, C., Rasko, J.E.J., 2008. Whither prometheus' liver? Greek myth and the science of regeneration. Ann. Intern. Med. 149, 421–426.

Raggi, C., Mousa, H.S., Correnti, M., Sica, A., Invernizzi, P., 2016. Cancer stem cells and tumor-associated macrophages: a roadmap for multitargeting strategies. Oncogene 35, 671–682.

Singh, A.M., Hamazaki, T., Hankowski, K.E., Terada, N., 2007. A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells 25, 2534–2542.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., Yamanaka, S., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131, 861–872.

Tamm, C., Galitó, S.P., Annerén, C., 2013. A Comparative Study of Protocols for Mouse Embryonic Stem Cell Culturing. PLOS ONE 8, e81156.

The Life of Dolly | Dolly the Sheep [WWW Document], n.d. URL http://dolly.roslin.ed.ac.uk/facts/the-life-of-dolly/index.html (accessed 5.22.18).

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., Lungjangwa, T., Imsoonthornruksa, S., Stelzer, Y.,

Rangarajan, S., D'Alessio, A., Zhang, J., Gao, Q., Dawlaty, M.M., Young, R.A., Gray, N.S., Jaenisch, R., 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15, 471–487.

Takara Bio Announces the Launch of Human iPSC-Derived Beta Cells for Drug Screening and Diabetes Research, n.d.

Mauro, A., Shafiq, S., Milhorat , A., 1970. Regeneration of striated muscle, and myogenesis .
van den Boom, V., Maat, H., Geugien, M., Rodríguez López, A., Sotoca, A.M., Jaques, J., Brouwers-Vos, A.Z., Fusetti, F., Groen, R.W.J., Yuan, H., Martens, A.C.M.,

Stunnenberg, H.G., Vellenga, E., Martens, J.H.A., Schuringa, J.J., 2016. Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Rep 14, 332–346.

Voigt, P., Tee, W.-W., Reinberg, D., 2013. A double take on bivalent promoters. Genes Dev. 27, 1318–1338.

Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J.D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., He, C., Santos, A.M., Helliwell, C., Pruneda-Paz, J.L., Kay, S.A., Lin, X., Cui, S., Garcia, M.F., Clarenz, O., Goodrich, J., Zhang, X., Austin, R.S., Bonasio, R., Wagner, D., 2017. Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. Nat. Genet. 49, 1546–1552.

Zhang, P., Torres, K., Liu, X., Liu, C., Pollock, R.E., 2016. An Overview of Chromatin-Regulating Proteins in Cells. Curr Protein Pept Sci 17, 401–410.

# Chapter 2: Mechanisms of *in vitro* Myogenesis of Human Induced Pluripotent Stem Cells, part 1

## Introduction

Human induced pluripotent stem cells have the potential to play an integral role in a number of personalized, regenerative medicine strategies. Chief among these are autologous cell-based therapies that could ameliorate compromised tissues, and patient-specific in vitro tissue- and disease models for drug testing and other diagnostics. Indeed, a limited number of hiPSC-based applications have started to make their way into clinical trials. However, despite successes in a few contexts, a major hurdle to the realization of the untapped potential of hiPSCs remains the reliable and robust differentiation to more terminal tissue types of choice.

Patient-specific hiPSC myogenic progenitors could be of great clinical utility, as there are unmet therapeutic needs for aging and injured muscle, as well for a number of myopathies, including Duchenne muscular dystrophy. Several strategies for the myogenic differentiation of human pluripotent stem cells (hPSCs) have been described in the literature. Embryoid body based differentiation, which consists of initial suspension of hPSCs for spontaneous differentiation into all three germ layers, followed by attachment culture and FACS purification for mesoderm progenitors has resulted in the derivation of a myogenic population that can give rise to PAX7 positive cells when transplanted in vivo. (Hwang, 2013; Kabra, 2015) However, FACS purification for mesoderm progenitors can be problematic, as the markers for sorting may not be specific for myogenic progenitors.

Another differentiation strategy relies on the ectopic over-expression of key transcriptions factors that regulate myogenesis. (Goudenege 2012) hiPSCs that were derived from Duchenne muscular dystrophy patients were differentiated with ectopic MyoD over-expression, delivered via lentivirus, and could be engineered with CRISPR-Cas9 in vitro to re-express dystrophin when transplanted in vivo. (Young, 2016). In addition, lentivirus-delivered, doxycycline-inducible Pax7 over-expression was used to differentiate hPSCs to form in vitro 3D tissue "bundles" that exhibited physiological responses to electrical and acetylcholine stimulation. (Rao, 2018). Although ectopic overexpression is an efficient method, the use of lentiviruses or adenoviruses could result in clinically unfavorable transgene introduction. Finally, several protocols for small molecule and growth factor directed monolayer myogenic differentiation report relatively efficient myogenic differentiation of hPSCs, and the resulting populations can then be further purified by FACS. Interestingly, these strategies result in the appearance of Pax7 positive cells, alongside multinucleated myotube-like cells, perhaps indicating the spontaneous creation of a satellite cell niche in vitro. (Borchin, 2013) More recently, transcriptomic studies of mouse (Chal, 2015) and human (Xi, 2017) in vivo somite development have helped to refine the in vitro cues for hPSC differentiation. Nevertheless, this strategy is limited by high variability, continuing problems with efficiency, and the long time required for differentiation, compared to ectopic overexpression.

As a result, the efficient, transgene-free differentiation of hiPSCs to the myogenic lineage continues to be an area of active research. Although protocols that successively improve the efficiency of myogenic differentiation have been published over the last five

years (Borchin, 2013; Chal, 2015; Xi, 2017), a mechanistic understanding of in vitro specification of human pluripotent cells to the myogenic lineage is lacking. To investigate the mechanisms of in vitro myogenesis of human induced pluripotent stem cells, we compared the longitudinal gene expression profiles of three hiPSC lines during the induction of transgene-free myogenic differentiation using a published protocol. (Chal, 2016) Longitudinal transcriptomic analyses of in vitro differentiation of hiPSCs could shed light on the nature of the temporally varying hierarchies of the key players, including transcription factors, of the gene regulatory networks that govern each successive stage of lineage specification towards terminal commitment.

Most transcriptomic studies of hPSC-derived, terminally specified lineages rely on FACS purification before sequencing, since there is better interpretation of gene expression data of a relatively homogenous population. Derivatives of all the germ layers have been studied in this way: neuronal progenitors (Rizzo, 2017), cardiac progenitors (Li, 2015), kidney progenitors (Kumar, 2015), and pancreatic beta cells (Huang, 2017), though transcriptomic studies of hPSC-derived paraxial mesoderm progeny are lacking. This strategy can make longitudinal study challenging, since the markers for sorting may change over the course of differentiation. Still, a few longitudinal transcriptomic studies of hPSC-derived cells focus on cardiogenesis (Tompkins, 2017) and neurogenesis (Li, 2017). Other transcriptomic studies have focused on understanding the pluripotent state, as in the reprogramming of somatic cells to hiPSCs (Wang, 2018), or on early commitment to an initial germ layer, within 48 hours from induction of differentiation (Huggins, 2017). On the other hand, transcriptomic studies of myogenesis have been carried out either in vivo,

to understand myogenesis in the context of regeneration in response to injury (Aguilar, 2015), or in vitro, in the C2C12 murine myoblast cell line that is already specified to the myogenic lineage (He, 2017). Transcriptomic studies of in vivo somite development (Chal, 2015; Xi, 2017) do not start at the pluripotent epiblast stage- which is considered the in vivo analogue to pluripotent cells in vitro (Nichols, 2009)- but rather at the tailbud. To our knowledge, this the first longitudinal transcriptomic study of hiPSCs that spans pluripotency to terminally specified myogenic progenitors.

We characterized the progression of myogenic differentiation in each of the three cell lines with immunofluorescent staining for multiple markers of early and late specification to the myogenic lineage, including Myf5, MyoG, Pax7 and myosin heavy chain. This characterization showed that one cell line had accelerated myogenic differentiation compared to the other cell lines; we used this temporal trend as a guide to identify the mechanisms that lead to more efficient myogenic differentiation compared to those process that retard/lead to off-target cell fates. Importantly, times series data allows us to trace back potential differential upstream lineage specification that may result in accelerated versus blunted myogenesis.

Furthermore, we take advantage of the inherent variabilities in hiPSC differentiation bias to guide our inquiry into mechanisms of robust in vitro myogenesis through our comparison of cell lines with differential myogenic commitment. Although by definition hiPSCs are pluripotent and can therefore give rise to all three germ layers (Pera, 2010), several studies have documented that different hPSC lines may preferentially give rise to a particular germ layer (Siller, 2016; Hu, 2016; Allegrucci, 2007). This bias in

differentiation tendency may arise from differential epigenetic priming of the key genes that regulate lineage specification (Bilic, 2011). In addition, some aspects of epigenetic memory may persist post reprogramming, and may influence subsequent differentiation (Nishinio, 2011). The comparison of cell lines with varying temporal dynamics of myogenic induction can allow for the identification of the key players and pathways that either promote or blunt in vitro myogenic differentiation of hiPSCs.

To accomplish this grouping together of genes that may function in the same processes, we assessed the broad trends in the longitudinal gene expression profiles between the cell lines by multiple methods. First, based on clustering of time points, we assigned each gene to an early, middle, or late stage of differentiation within each cell line separately and compared the gene expression of each cell line in each stage of differentiation. Finally, we performed hierarchical clustering to the overlapping or complementary gene groups between the cell lines in each stage of differentiation to identify modules of genes with significant biologically relevant functional enrichment.

## Results

### Temporal appearance and extent of commitment of myogenic progenitors varies between hiPSC lines

We induced transgene-free myogenic differentiation in three hiPSC lines according to a published protocol (Chal, 2016). (Figure 2.1) All hiPSC lines were derived from peripheral blood mononuclear cells from healthy individuals by ectopic overexpression of the Yamanaka factors, delivered via Sendai virus. Cells were plated at low density as undifferentiated single cells at day zero; in the subsequent thirty days of differentiation we

observed that the hiPSC line L-EPCC3 (L) had earlier and more widespread appearance of markers of terminal myogenic specification than the other two hiPSC lines: SCVI15 (S) and TL (T).

We characterized the temporal progression of myogenic specification in each of the three cell lines with immunofluorescent staining for multiple markers of myogenic differentiation. (Figure 2.2) The transcription factor Myf5 was detectable by day 6 in all three cell lines, with expression peaking around day 12-16. This appears to be consistent with reports of Myf5 expression in embryonic somatic dermomyotome (Buchberger, 2003), from which myogenic progenitors delaminate into the limb buds. However, its expression has also been reported in brown adipocytes (Shan, 2013), and Myf5 transcripts, but not protein, were also detected in the developing central nervous system (Daubas, 2000). Thus, Myf5 expression is indicative of, but not specific to, the myogenic lineage.
In the later time points, we observed time-dependent expression of key markers specific to myogenesis. By day 19, all cell lines began to express desmin, a nonspecific cytoskeletal marker found in skeletal muscle and other cell types. By day 25, a difference in the expression of markers of skeletal muscle was apparent, as the L line began to express myosin heavy chain (MF20), MyoG a key MRF for terminal myogenic specification, as well as numerous Pax7 positive cells, while the other two cell lines did not. Since the Pax7 positive cells arose in spatial and temporal coincidence with Mf20 and MyoG positive cells, it is likely that they represent a satellite cell-like population. By day 30, the S line began to express Mf20 positive cells, while the T cell line only had limited subpopulations

of cells expressing Mf20. By contrast, the L cell line had widespread Mf20, MyoG and Pax7 expressing cells by day 30.

Whole transcriptomic profiling of three hiPSC lines shows increasingly divergent gene expression with increasing time of differentiation

We sought to understand the mechanistic underpinnings of robust versus blunted in vitro myogenic differentiation by comparing the longitudinal whole-transcriptomic gene expression of the three hiPSC lines. We carried out RNA sequencing on 9 sequential time points from each of the three cell lines, spanning day 0 pre-induction undifferentiated cells to day 30. Figure 2.3 summarizes the experimental design and computational pipeline for read alignment and differential expression analysis. Raw reads were assessed for quality using FastQC- all samples had high mean quality scores and normally distributed GC content, centered around 50%. (Figure 2.4). Across all samples in all cell lines, an average of 86.8% of raw reads aligned to the transcriptome consisting of both protein-coding cDNA and non-coding RNA. (Figure 2.5).

We used the TPM normalized gene counts from alignment to the transcriptome to calculate pairwise Pearson correlation between all samples for all time points belonging to each cell line separately. Figure 2.6A shows a representative correlation heat map for one cell line. As expected, biological replicates from the same time point are most similar to each other followed by samples that are closest in time. Samples that are farthest in time from each other tend to be most dissimilar, as expected during the progression of differentiation.

Similarly, we also calculated pairwise Pearson correlation between the different cell lines, for samples at the same time point (Figure 2.6B). At undifferentiated day 0, all replicates from all cell lines have a relatively high degree of correlation. With increasing time of differentiation, although replicates from the same cell line have high correlation, samples from different cell lines have lower correlation. In particular, the L line, which exhibited the most robust myogenic differentiation, appears to diverge from the S, T lines relatively early in differentiation, and this low correlation between the L line and the other two is especially marked by day 30 late differentiation.

The broad findings from differential gene expression analysis are likewise consistent with this trend. (Figure 2.7) We analyzed each cell line separately, with each time point normalized with respect to the day 0 sample for its respective cell line. After applying log2 fold change and adjusted p-value thresholds of 1 and 0.005, respectively, there were 16,588 differentially expressed genes across all time points and cell lines, out of a total transcriptome consisting of 56,864 transcripts. As expected, the number of differentially expressed genes (DEGs) increases with time of differentiation for all cell lines. The number of DEGs for the L cell line is greater than the number of DEGs in the other two cell lines at each time point, and this discrepancy increases with differentiation. Furthermore, hierarchical clustering of the time points from all cell lines, with respect to differential gene expression, generally grouped “early”, “middle” and “late” time points together. The resulting dendrogram (Figure 2.8) shows day 3 samples from all cell lines closely clustered together, along with day 6 and 8 points from the S and T lines only. Interestingly, these early points were closer to days 12, 16 and 19 from the S and T lines,

which perhaps reflects the blunted or slowed differentiation observed in those cell lines compared to the L line. Days 25 and 30 from the S and T lines also clustered closely together. By contrast, the time points for the L line clustered away from the other two cell lines: days 6, 8, 12, and 16 from the L line clustered together, closer to the early and middle time points from the other cell lines. However, days 19, 25, and 30 from the L line clustered together, away from all the other time points. Taken together, the data indicates that the gene expression profiles of the three cell lines can be divided into "early", "middle" and "late" stages of differentiation. Moreover, the gene expression profiles of the three cell lines are somewhat similar in early time points, but the cell line with more robust myogenesis appears to increasingly diverge from the other two by the middle and late stages of differentiation.

Hierarchical clustering of overlapping genes between cell lines results in gene modules with biologically relevant functional enrichment

The large fraction of the total transcriptome that is differentially expressed across all time points and all cell lines may reflect heterogeneous cell populations that arise in the course of in vitro, transgene-free differentiation of hiPSCs. Therefore, it was important to segregate genes into groups that may function in concert to affect a particular biological process. To achieve this, we first binned genes into groups that reflect the early, middle, and late stages of lineage commitment for each cell line separately by computing a "center of mass"- type metric for each gene. (Figure 2.9). The cut-off values for whether a gene was classified as early, middle, or late were determined by the local minima of the trimodal

histograms of the counts of the “center of mass” metric. After binning the genes for each cell line separately, we next compared which genes were active in each early, middle, or late differentiation stage between the cell lines. Figure 2.10 summarizes the numbers of genes that are common to or different from each cell line in each differentiation stage. Since each cell line was treated separately, numerous genes are counted more than once, which may reflect variations in timing of differentiation-related processes between the cell lines. Finally, we performed hierarchical clustering within each intersecting and complementary set of genes at each of the early, middle, and late stages. We used a dynamic tree-cutting algorithm published by Horvath et al. for optimal gene module detection. Figures 2.11-2.14 show examples of two gene modules derived in this way. The gene module in figure 2.11 contains several key transcription factors involved in mesoderm specification, including brachyury (T), TBX6, MIXL1, and MSGN1. Significantly enriched terms (Fig 2.12) are related to endoderm, mesoderm and neural crest differentiation, as are biological processes related to embryonic axis specification and somitogenesis. Similarly the gene module in figure 2.13 contains a set of genes that is highly enriched for processes relating to terminal myogenic specification, including striated muscle contraction, hypertrophy, muscle development, and satellite cell differentiation (Fig 2.14). The MRFs MYOD1 and MYOG are present in this group, as well as numerous myosin heavy chains and a component of the acetylcholine receptor (CHRNA1). Interestingly, a known long noncoding RNA, LINCMD1 is also present in this group; LINCMD1 promotes myogenic differentiation by acting as a sponge for inhibitory miRNAs (Cesana, 2011).

Identification of unannotated transcripts that cluster closely with key MRFs and other TFs that regulate myogenesis, myosin heavy chains, and LINCMD1

We identified four unannotated transcripts (Figure 2.15) that appear in the cluster (Figures 2.13, 2.14) that is highly enriched for genes related to myogenesis. Of these transcripts, encode classifies two as putative lincRNAs, one as a retained intron, and one as a putative protein coding transcript. We also tabulated the genomic neighborhood of each transcript to check for overlap of other transcripts, as this may give insight into possible function. One transcript overlapped the gene EMC10, which was not differentially expressed in our RNA seq data. Of note, one transcript (AC083902.2, putative protein-coding) was also found to have numerous orthologues in different species indicating possible evolution conservation. We next used the genome browser (UCSC Genome Browser, GRCh37/hg19) with added tracks showing CHiP seq data from primary human myoblasts for several histone modifications (Bradley Bernstein Lab, Broad). We also used the tissue specific transcript expression database (GTEx RNA-seq) that is also integrated with the genome browser. We used LINCMD1 as a "positive control" (Figure 2.16), since it is a known lncRNA that promotes myogenic differentiation. CHiP-seq tracks show enriched peaks for activating histone marks upstream of LINCMD1, and tissue specific gene expression shows expression specifically in skeletal muscle, out of 53 possible tissue types. Similarly, for the unannotated AC083902.2 transcript, we noted activating H3K4me3 peaks upstream of the transcript as well as skeletal muscle specific expression. Other transcripts showed either no histone mark enrichment, or did not have skeletal

muscle specific expression (data not shown). This indicates that AC083902.2 may be a previously unknown, putative protein-coding gene involved in myogenic differentiation.

**Discussion**

We investigated the mechanisms of in vitro myogenesis of human induced pluripotent stem cells with longitudinal gene expression profiles of three hiPSC lines during the induction of transgene-free myogenic differentiation. We characterized time-dependent commitment to the myogenic lineage of the three hiPSC lines by immunofluorescent staining for multiple transcription factors and cytoskeletal proteins that are markers of myogenic specification. (Figure 2.2) Myf5, a transcription factor known to be an early regulator of dermomyotome commitment, was expressed by mid-differentiation in all three cell lines; however, since Myf5 is not specific to myogenic progenitors, this may indicate a different dermomyotome-derivative fate (Shan, 2013), in addition to potential myogenic progenitors. In the later time points, we observed that the L cell line expressed Mf20, Myog, and Pax7 by day 25, and expression of these markers increased by day 30 of differentiation. By contrast, the S and T lines had relatively few cells that expressed Mf20 only, with no MyoG or Pax7 expression detectable by immunofluorescence, at day 30. This indicates that the L line exhibited more robust myogenesis in 30 days of differentiation compared to the other two cell lines. In addition, the appearance of Pax7 positive cells in the L line, in temporal and spatial proximity to Myog and Mf20 positive cells could imply the spontaneous creation of a niche environment for satellite-like cells in vitro, which has been previously reported (Borchin, 2013, Chal, 2015) during in vitro myogenic differentiation of human pluripotent stem cells.

The broad trends in the data correspond well with the phenotypic characterization. Calculation of pairwise Pearson correlation using TPM normalized gene counts of replicates from all cell lines shows that correlation between cell lines decreases with increasing time of differentiation. (Figure 2.6) The L line, which had robust myogenesis, diverges significantly from the other two lines in the late time points, in agreement with the staining data. As a practical consideration, we note that, due to the divergent gene expression patterns between cell lines, it may be problematic to treat replicates from different cell lines as biological replicates, even at the same time point. Therefore, for differential gene expression analysis, we treated each cell line separately and normalized each time point to its respective day zero from the same cell line. Hierarchical clustering of the time points with respect to differential gene expression reinforces this increasing divergence between the cell lines; although the day 3 time points of all three lines cluster together, by day 6 the L line clusters separately from the S and T lines. Days 19, 25, and 30 from the L line especially cluster away from all the other time points, which may reflect the processes that give rise to terminally specified myogenic progenitors and Pax7 positive satellite-like cells that arise in the L line. (Figure 2.8)

It is possible that longer-term culture of the other two cell lines could result in the eventual appearance of MyoG and Pax7 positive cells, indicating delayed, as opposed to less efficient, myogenesis. Delayed myogenic differentiation in one cell line versus another could reflect the diverse embryonic origins of skeletal muscle in vivo. For example, the musculature of the limbs versus extraocular and head muscles have separate embryonic origins (somitic mesoderm versus head mesenchyme) and even have varying core

transcriptional regulatory mechanisms that determine specification (Braun, 2011). In addition, it is known that paracrine signaling from surrounding tissues affects myogenic specification at numerous stages of development. (Schoenwolf, 2015) Epiblast priming and primitive streak development are influenced by numerous signals from extraembryonic tissue as well from the hypoblast. (Mesnard, 2011) Later in embryonic development, growth factor gradients and paracrine signals from the developing notochord and lateral plate mesoderm determine the rate of somite maturation. (Schoenwolf, 2015) Finally, it has been shown that patterning of the myotome depends on signals from motor neurons. (Borchin, 2013) As a result, there are numerous possibilities for variations in auxiliary cell populations that could account for time shifts in the appearance of myogenic progenitors in vitro.

On the other hand, it is also possible that the relatively poor myogenesis at thirty days of differentiation reflects poor specification to paraxial mesoderm at the outset of differentiation that cannot be rectified with additional culture time, as the cells may be committed to another lineage altogether. Pluripotent hiPSCs under standard culture conditions bear epigenetic memory that could lead to variation in the priming of key lineage specifying genes, and this may bias lineage specification upon induction of differentiation.

Indeed, given the heterogeneous nature of in vitro hiPSC differentiation, it is likely that these processes may occur to various degrees simultaneously in the different cell lines to determine the overall efficiency of myogenic differentiation. However, several trends in the transcriptomic data lend support for the idea that differences in initial germ layer

specification could be responsible for the observed downstream differences in myogenic specification. First, about one third of the total transcriptome- that is expanded with the addition of non-coding RNAs in addition to the usual protein coding cDNA transcripts- is differentially expressed. (Figure 2.7) The large fraction of differentially expressed genes, across all time points and all cell lines, implies a heterogeneous differentiation in which multiple cell lineages that perhaps derive from different germ layers are present. Second, since the cells are cultured in the presence of high concentration CHIR up to day 6, beta catenin-mediated signaling should be highly active at this point. (Figure 2.1) The hierarchical clustering of time points with respect to differential gene expression shows that the L cell line clusters away from the S and T cell lines as early as day 6 onwards, indicating that the cell lines diverge in their gene expression, even in the presence of small-molecule mediated canonical Wnt activation. Wnt activation is the initial step to induce, not only paraxial mesoderm, but also cardiac mesoderm and endoderm as well, illustrating a possible route to divergent germ layer specification and corresponding gene expression.

In summary, our experimental strategy addresses several challenges: hiPSC differentiation varies from cell-line to cell-line, and is known to be heterogeneous in terms of both the cell populations that arise during differentiation, as well as the temporal course of differentiation. The use of longitudinal data from three different hiPSC lines makes allowances for both these factors, since we are able to create time-dependent gene expression profiles across all three cell lines. In order to separate out relevant processes, we reasoned that genes that played a role together in a specific biological process were likely to have similar patterns of expression across all three cell lines. In this way, we were

able to “triangulate” gene modules with biologically relevant functional enrichment. (Figures 2.11, 2.12) Furthermore, a main utility of longitudinal whole transcriptome data is that it allows for the establishment of relationships between upstream genes and their downstream targets, including expression of transcription factors and their targets, and time-evolution of signaling pathway components. Combining this information from all three cell lines will allow us to compare and contrast groups of genes and their temporal expression patterns in multiple hiPSC lines thus isolating cell-line idiosyncratic gene expression from key lineage commitment gene expression patterns. In this, we are guided by phenotypic characterization of the progression of myogenic differentiation in all three cell lines, so we can link temporal gene expression to the trends observed by immunofluorescent staining. Finally, we identify and characterize a promising unannotated, putative protein-coding transcript that may be involved in myogenic differentiation.

In the next aim, we look more closely at the expression profiles of key transcription factors and their upstream and downstream partners, epigenetic modifiers, and key components of gene regulatory networks that could explain differences in myogenic specification in the three hiPSC lines.

## Figures

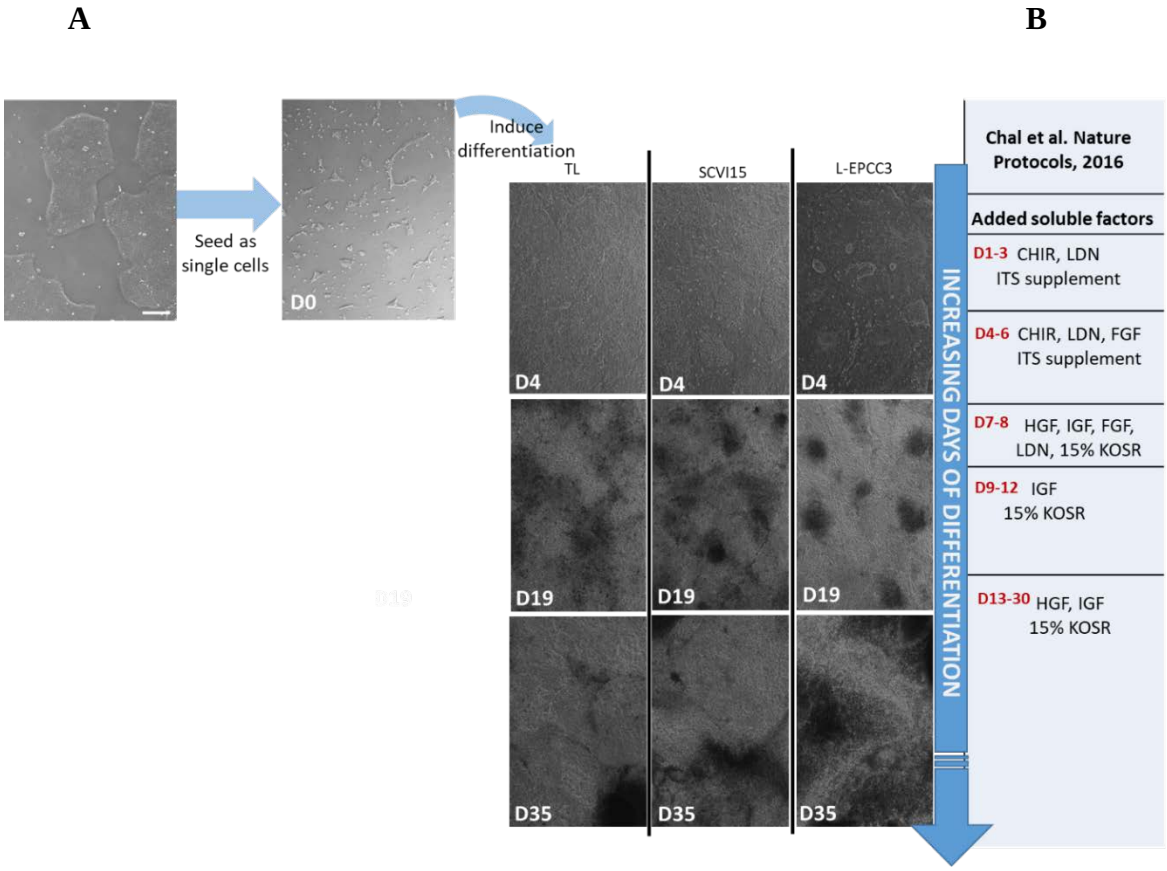



**Figure 2.1**: **Myogenic differentiation of three human hiPSC lines according to a published protocol**

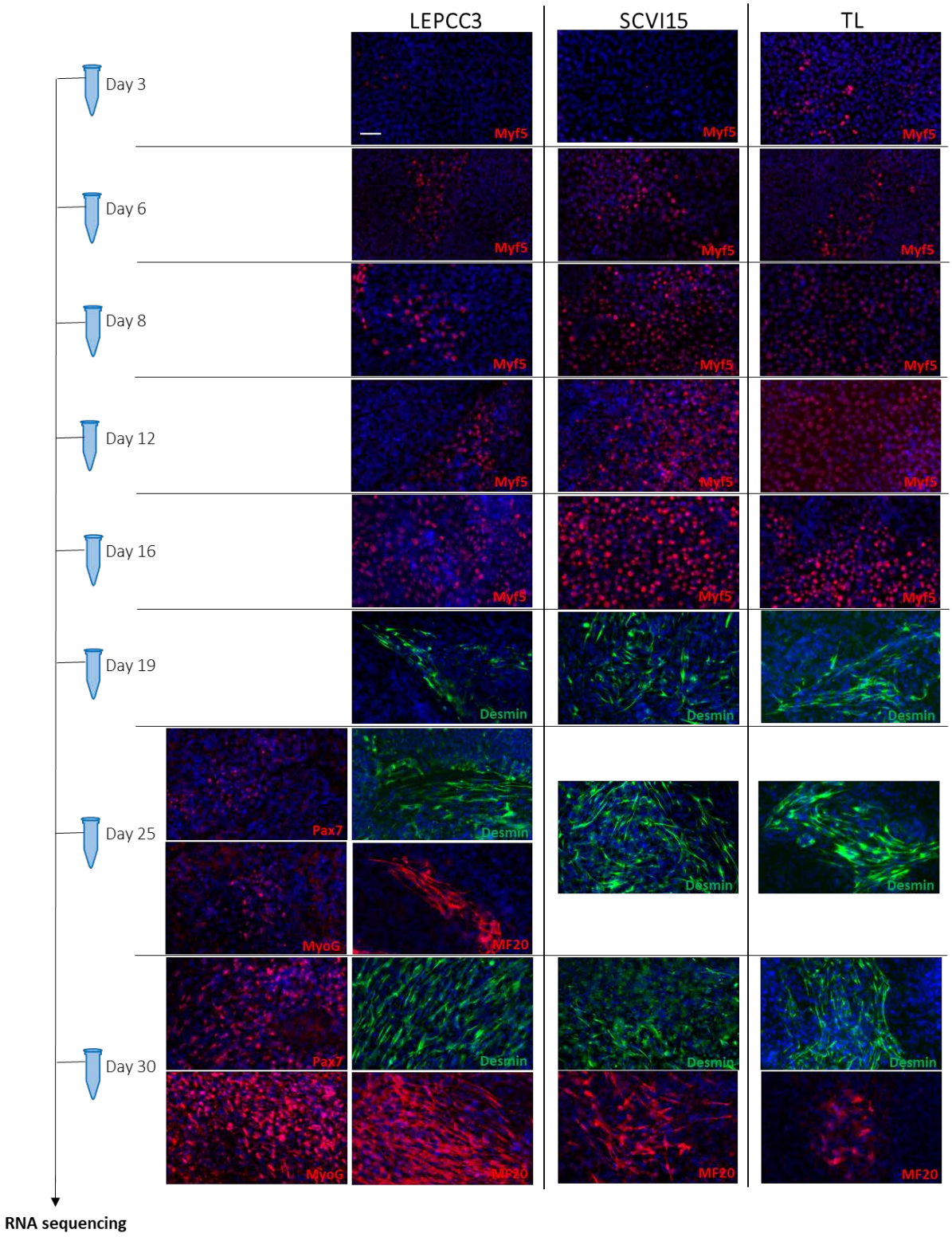
A. Brightfield images depict 35 days of myogenic differentiation of three hiPSC lines using a published protocol by Chal et al. Cells were maintained as pluripotent colonies and seeded as single cells prior to induction of differentiation. The cells become confluent by day 4 and continue to proliferate, forming dense clusters that appear as diffuse dark areas under brightfield microscopy. There is some cell-line to cell-line variability, but overall all three lines show a similar trend across 35 days under brightfield microscopy.

B. We used a published protocol by Chal et al to induce myogenic differentiation with the listed small molecule and growth factor supplements at various stages of differentiation.

**Figure 2.2: Characterization of temporal appearance and extent of commitment of myogenic progenitors during myogenic differentiation of hiPSCs shows accelerated myogenic specification of LEPCC3 cell line**

Three hiPSC lines (LEPCC3, SCVI15, and TL) show positive immunofluorescent staining for the transcription factor Myf5 early in differentiation. By day 19, all three lines show desmin expression. By day 25, only the LEPCC3 line has cells that are positive for Mf20, MyoG and Pax7. By day 30, the LEPCC3 line has numerous cells positive for Mf20, MyoG, and Pax7. The SCVI15 and TL lines have few cells positive for Mf20 by day 30- SCVI15 has more positive cells than TL- but neither have MyoG or Pax7 positive cells detectable by immunofluorescent staining.

Eppendorf symbols to the left represent experimental design for collection of samples for RNA sequencing that capture the time shift in expression of myogenic markers between the cell lines. Two biological replicates were collected from each of the cell lines for RNA sequencing at each of the time points: day 0 (undifferentiated), day3, day 6, day 8, day 12, day 16, day 19, day 25, and day 30, for a total of 54 samples.


LEPCC3
SCVI15
TL
Day 3
Day 6
Day 8
Day 12
Day 16
Day 19
Day 25
Day 30
Myf5
Desmin
Pax7
MyoG
MF20
RNA sequencing

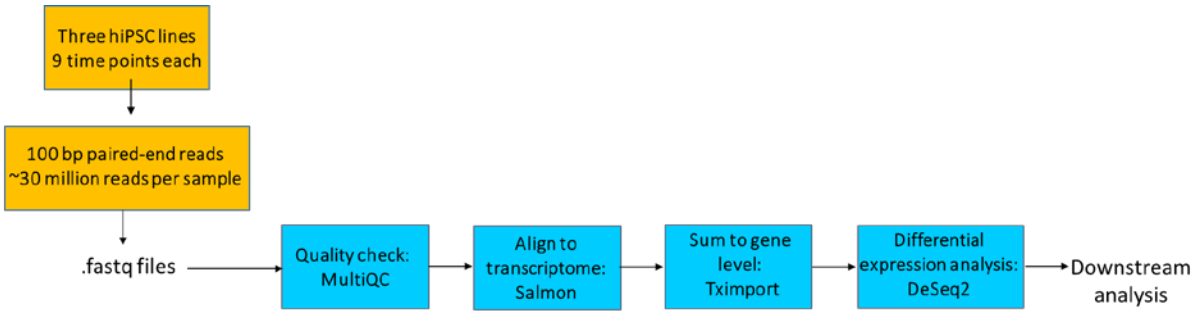



**Figure 2.3: Schematic of computational pipeline for alignment of raw reads to transcriptome and differential expression analysis**

Raw reads were assessed for quality and aligned to a transcriptome consisting of protein coding cDNA and noncoding RNA transcripts. Each time point was normalized with respect to the day 0 undifferentiated sample from its respective cell line; differential expression was calculated for each cell line separately.


**Figure 2.4: Raw read quality assessed with FastQC**
FastQC was used to assess the the read quality of each sample, and outputs for multiple samples assessments were compiled with MultiQC.

A. The mean quality score is high across the length of each read
B. The majority of reads have a mean sequence quality score above 30
C. The %GC content is normally distributed around 50%
D. The per base n content is less than 1% for all positions across each read

A

Representative proportion of reads aligned

| Sample Name | % Aligned | M aligned |
|---|---|---|
| LEPCC3_D0_1 | 87.90% | 26.2 |
| LEPCC3_D0_2 | 86.20% | 27.4 |
| LEPCC3_D3_1 | 88.70% | 22.6 |
| LEPCC3_D3_2 | 90.20% | 27.3 |
| LEPCC3_D6_1 | 87.30% | 31.7 |
| LEPCC3_D6_2 | 88.30% | 34.5 |
| LEPCC3_D8_1 | 86.80% | 29.5 |
| LEPCC3_D8_2 | 85.80% | 33.2 |
| LEPCC3_D12_1 | 87.60% | 30.2 |
| LEPCC3_D12_2 | 88.30% | 25.8 |
| LEPCC3_D16_1 | 86.00% | 30.1 |
| LEPCC3_D16_2 | 86.30% | 25 |
| LEPCC3_D19_1 | 84.70% | 28.3 |
| LEPCC3_D19_2 | 86.70% | 28.4 |
| LEPCC3_D25_1 | 86.70% | 26.2 |
| LEPCC3_D25_2 | 85.80% | 26.5 |
| LEPCC3_D30_1 | 83.90% | 28.3 |
| LEPCC3_D30_2 | 85.90% | 32.6 |

B

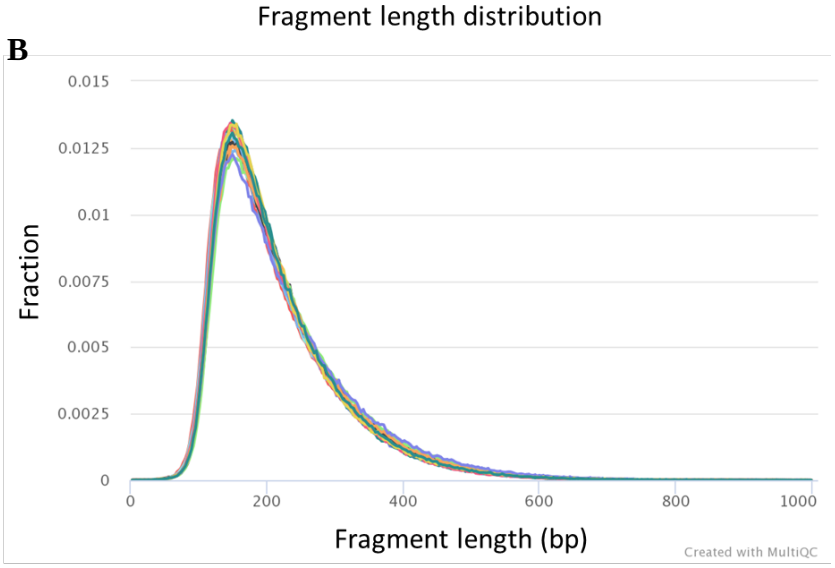



**Figure 2.5: Read alignment to transcriptome**

A. Representative fraction of reads aligned for one cell line. The average percentage of reads aligned was 86.8 ±1.1% and the average number of reads per sample was 29.1 ± 3.0 million reads.

B. Fragment length distribution for a representative group of samples shows fragment length distribution is smooth and peaks at 200 base pairs.

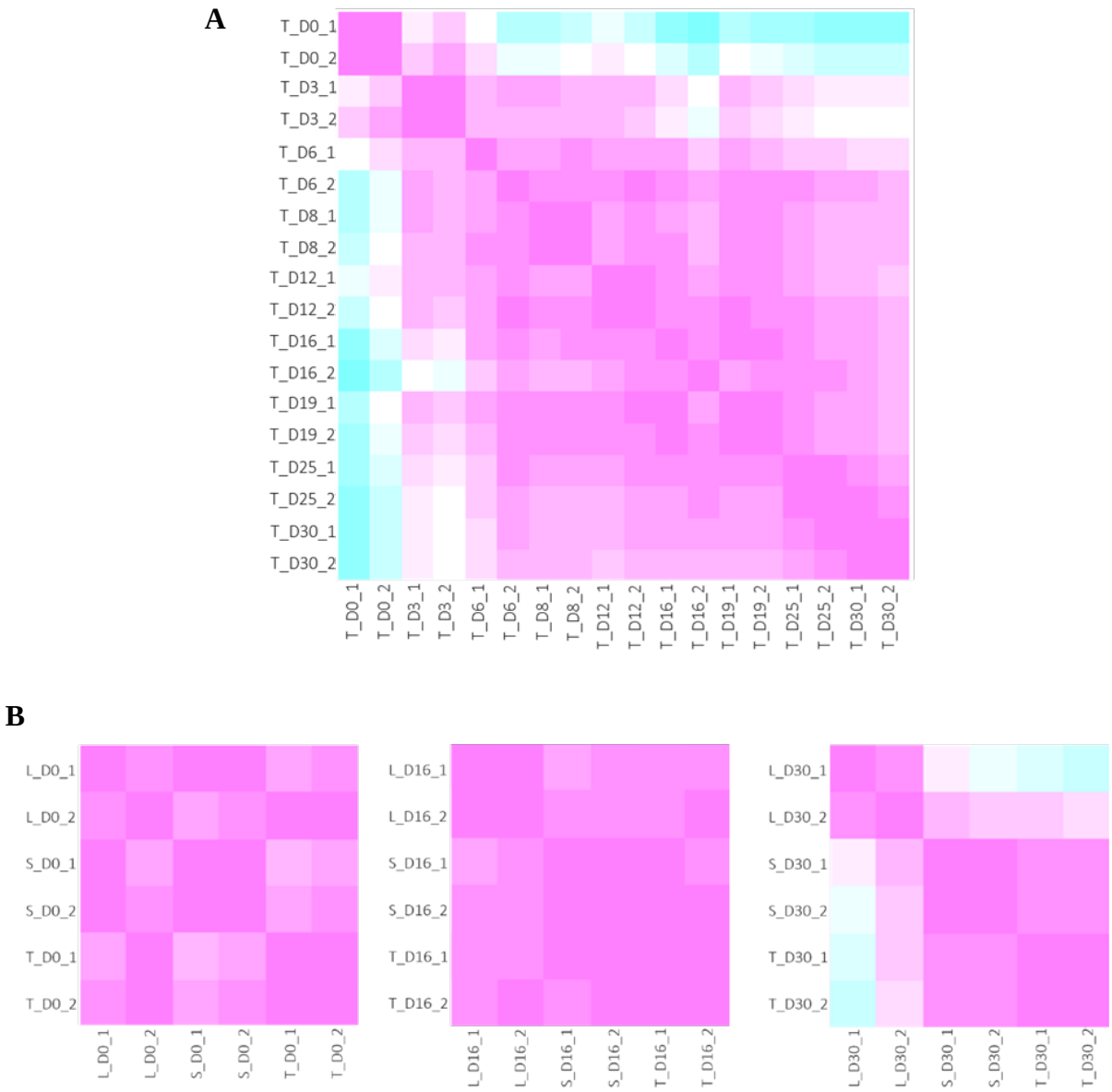



**Figure 2.6 Pearson correlation for replicates across time and across different cell lines**

A. Pairwise Pearson correlation for all time points and replicates for TPM normalized gene counts for a representative cell line. Biological replicates have high correlation with each other. Adjacent time points have higher correlation than distant time points.

B. Each heat map represents pairwise Pearson correlation calculated for TPM normalized gene counts between all replicates from all cell lines. At day 0, correlation is high between all the cell lines. At day 16, correlation is still relatively high, but the SCVI15 and TL lines are more similar. By day 30, replicates from the same cell line have high correlation, but L line has low correlation with the other two lines.


**Figure 2.7: Number of differentially expressed genes with respect to time and cell line**

All three cell lines had as increasing number of differentially expressed genes with respect to time; each time point is normalized with respect to its respective undifferentiated day 0. The L cell line has more differentially expressed genes at each time point compared to the other two cell lines, and this discrepancy increases with increasing time of differentiation. By contrast, the number of differentially expressed genes are more similar in the S and T lines.

.

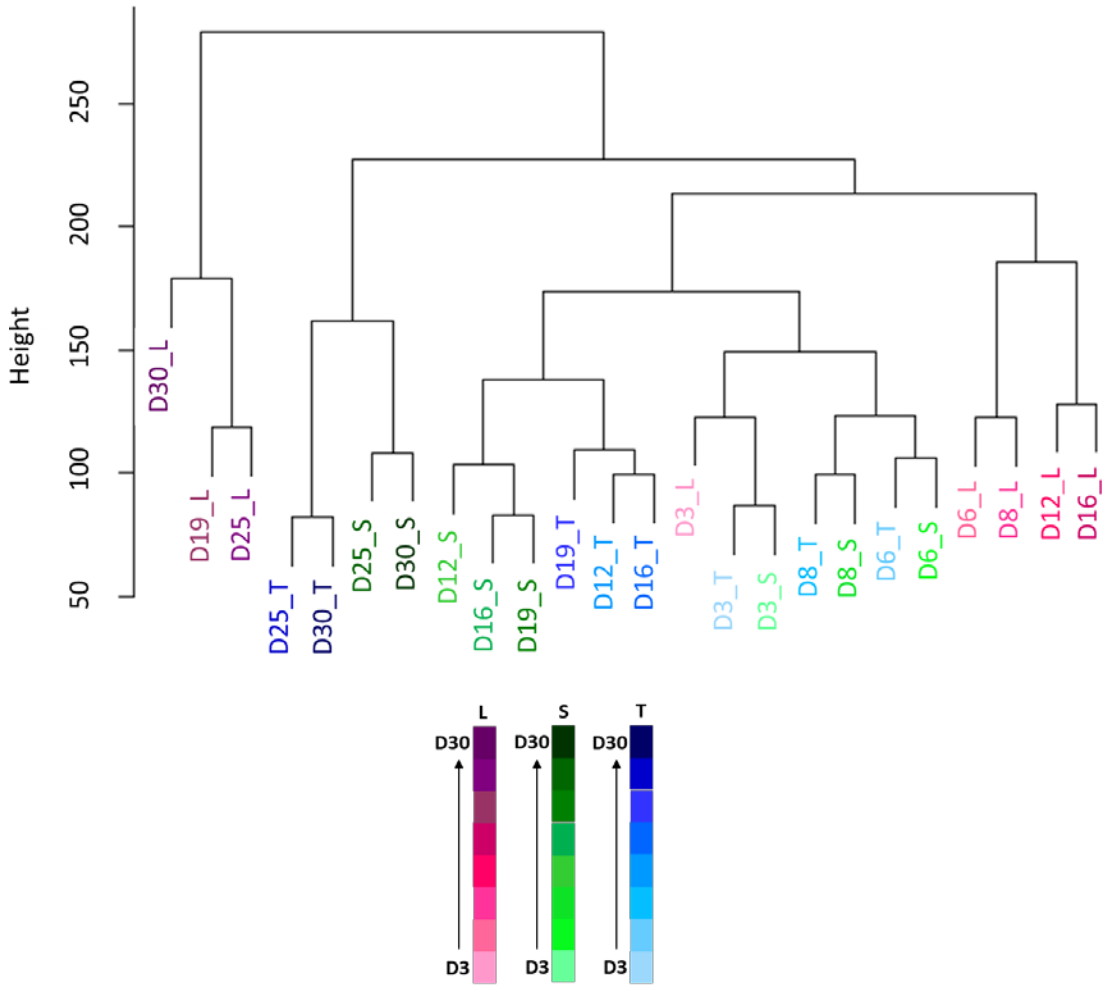



**Figure 2.8 Hierarchical clustering of time points from all cell lines**
Hierarchical clustering with respect to differential expression gene expression was performed for all time points from all cell lines. Day 3 time points from all cell lines clustered together, in proximity to the day 6 and day 8 samples for the S and T lines. Days 12, 16, and 19 from the S and T lines are likewise clustered together, closest to the early time points. Days 6, 8, 12, 16 for the L line are clustered together; Days 25 and 30 from the S and T lines are clustered together. Finally, days 19, 25, and 30 for the L line are clustered together, away from all the other time points.
.

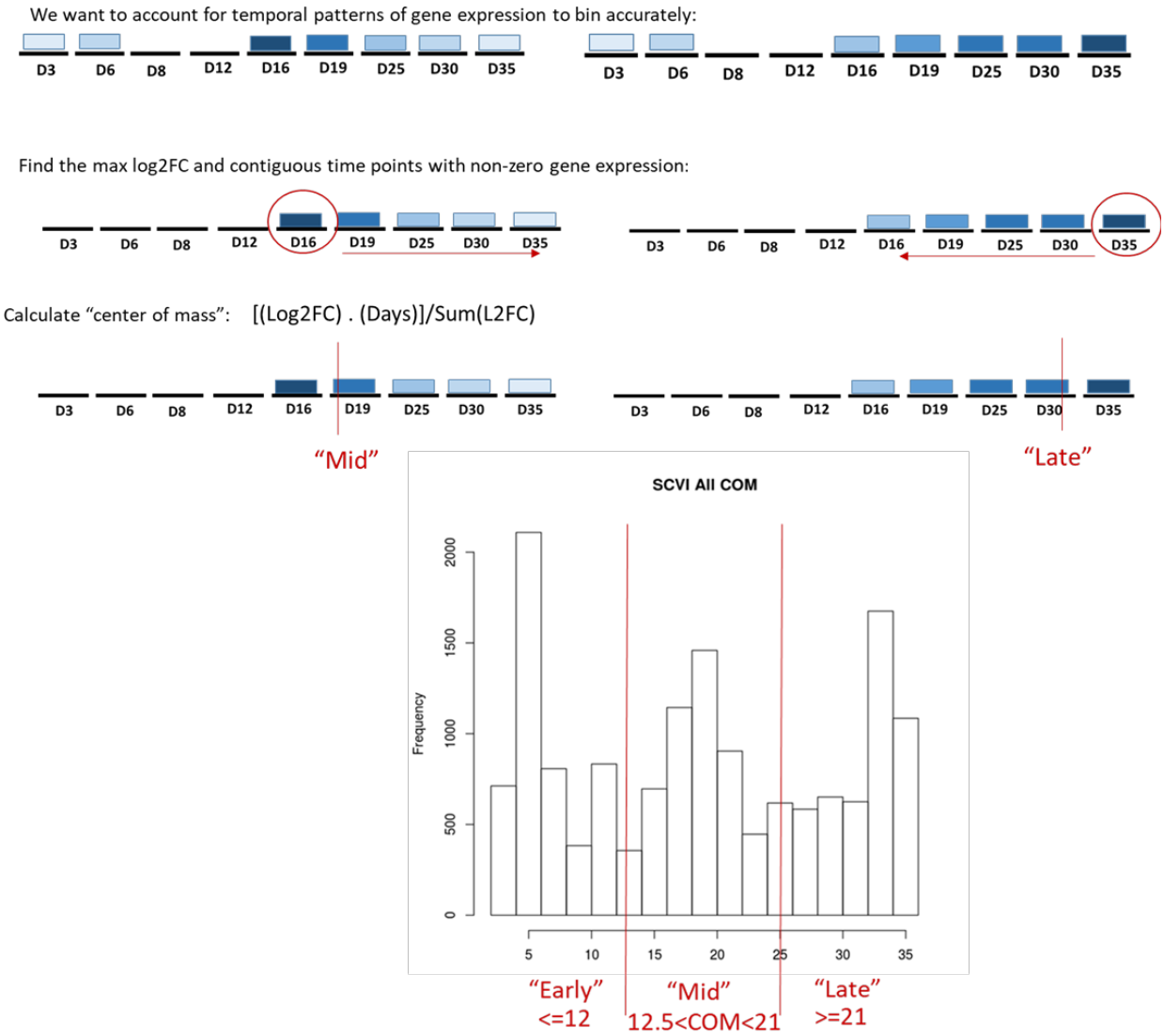



**Figure 2.9: Schematic for "center of mass" type calculation to bin genes into early, middle, or late stage of differentiation**

A. This schematic details the strategy used to calculate a metric that reflects the temporal expression pattern for each gene. Each cell line was treated separately

B. A representative histogram for one cell line of the counts of the "center of mass" type metric for each gene, as calculated in (A). Each cell line had a trimodal histogram, similar to the one above. Cutoffs for early, middle, and late were chosen according to the minima of the histogram.

.

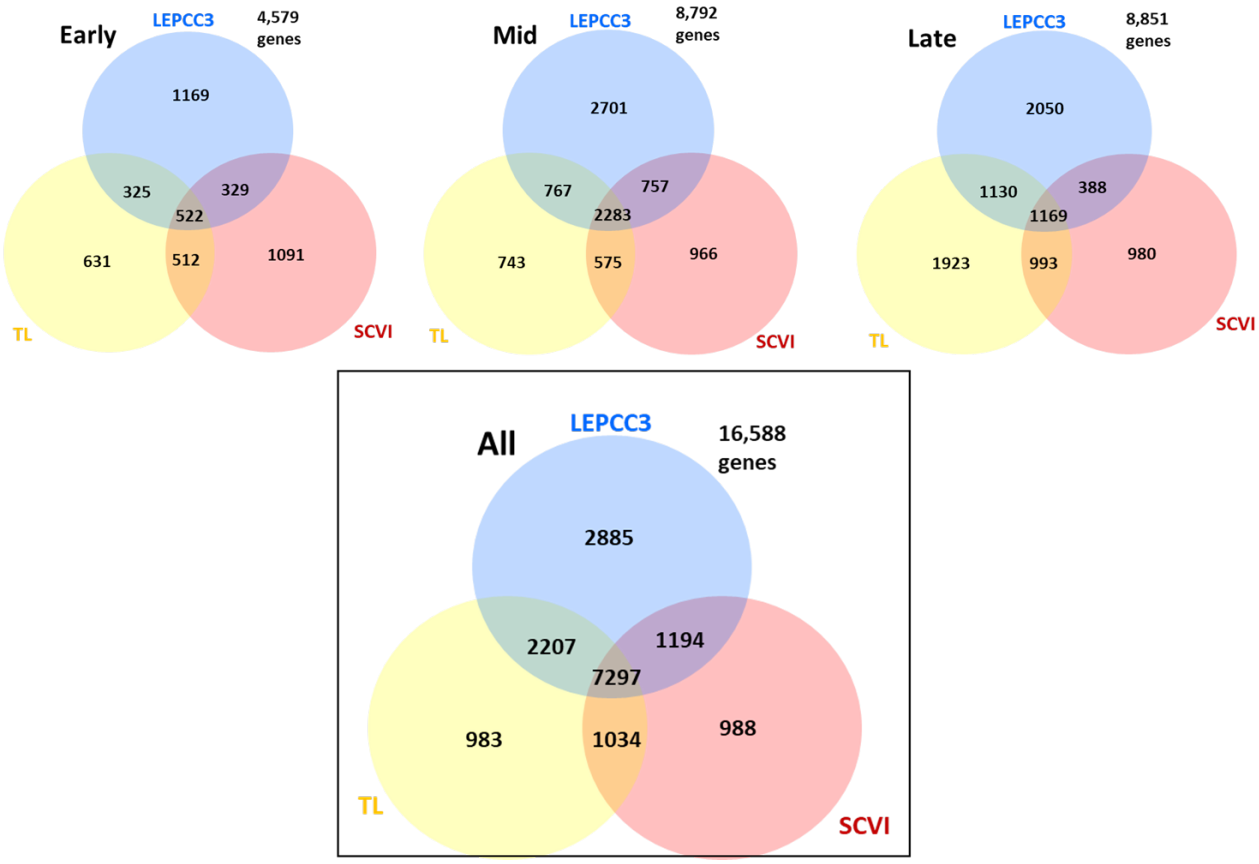


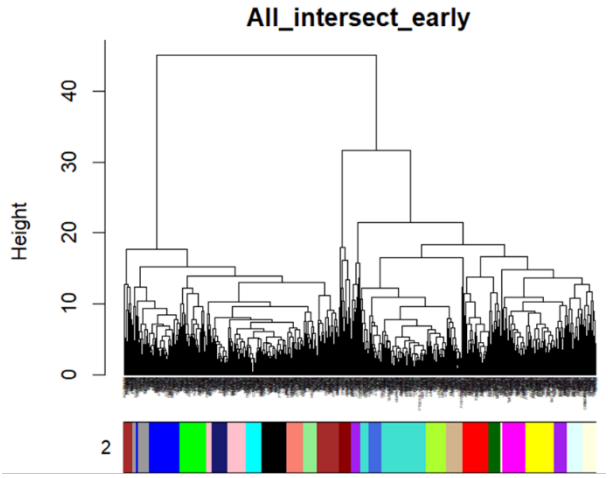

**Figure 2.10: Overlapping gene lists between each of the cell lines, in each of the early, middle, and late stages of differentiation**

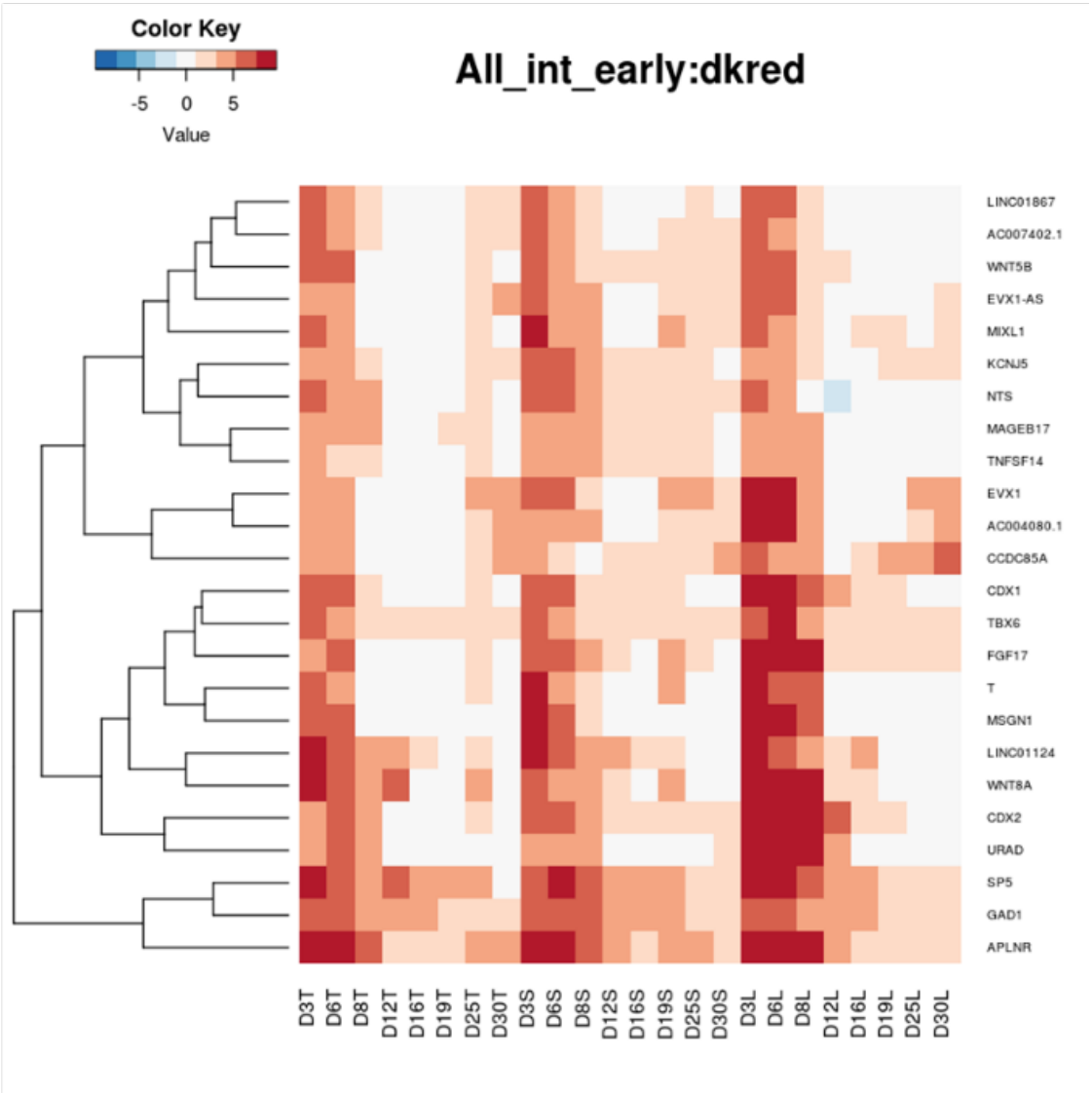
Triple Venn diagrams between each of the three cell lines for gene counts in each of the early, middle, and late stages of differentiation, as well as for all time points (boxed). The sum of the total genes in each of the categories is greater than the total number of genes due to differences in temporal expression of many genes in each cell line, leading to double or triple counting.

**Figure 2.11: Example of module for genes that were upregulated early in all three cell lines**

C. Hierarchical clustering of all the "early" genes that overlap in all three cell lines. The colored bar at the bottom represents the separate modules detected by the dynamic tree-cutting algorithm. (Langfelder, 2008)

D. Heat map of the "darkred" group as detected above. Genes have a distinct expression pattern across all three cell lines, and this group contains several well-known genes related to mesendoderm specification: T (brachyury), TBX6, CDX1, CDX2, MSGN1, MIXL1.


All_intersect_early
Height
0
10
20
30
40
2


Color Key
-5
0
5
Value
All_int_early:dkred
LINC01867
AC007402.1
WNT5B
EVX1-AS
MIXL1
KCNJ5
NTS
MAGEB17
TNFSF14
EVX1
AC004080.1
CCDC85A
CDX1
TBX6
FGF17
T
MSGN1
LINC01124
WNT8A
CDX2
URAD
SP5
GAD1
APLNR
D3T
D6T
D8T
D12T
D16T
D19T
D25T
D30T
D3S
D6S
D8S
D12S
D16S
D19S
D25S
D30S
D3L
D6L
D8L
D12L
D16L
D19L
D25L
D30L

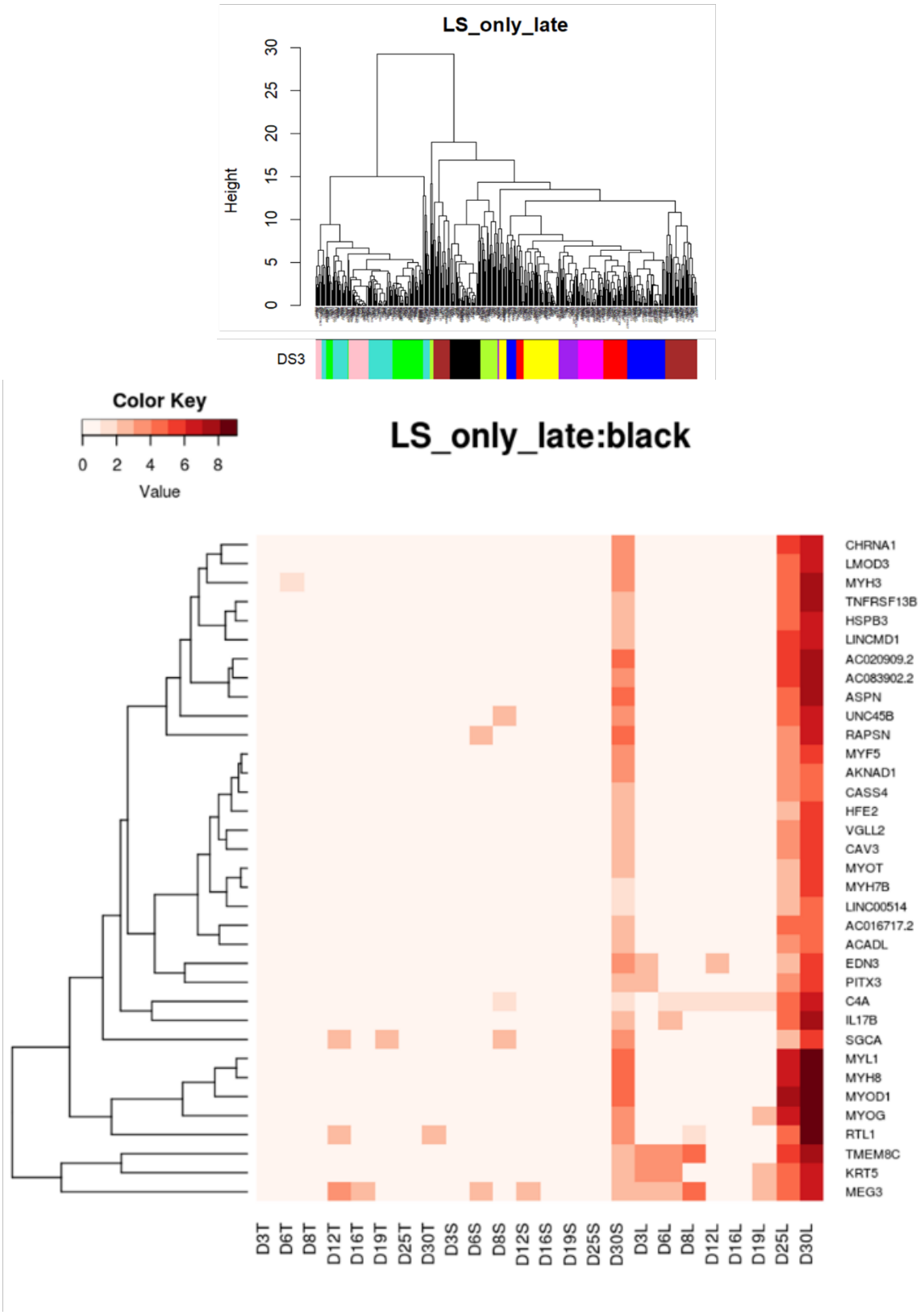
| Wikipathways 2016 | p-value |
|---|---|
| Cardiac Progenitor Differentiation_Homo sapiens_WP2406 | 0.001539 |
| Endoderm Differentiation_Homo sapiens_WP2853 | 0.002741 |
| Mesodermal Commitment Pathway_Homo sapiens_WP2857 | 0.003134 |
| Neural Crest Differentiation_Homo sapiens_WP2064 | 0.005461 |
| Adipogenesis_Homo sapiens_WP236 | 0.008895 |
| Gene regulatory network modelling somitogenesis _Homo sapiens_WP2854 | 0.01204 |

| GO Biological Process 2017b | p-value |
|---|---|
| anterior/posterior axis specification, embryo (GO:0008595) | 7.777e-7 |
| digestive tract mesoderm development (GO:0007502) | 0.000004601 |
| regulation of Notch signaling pathway involved in somitogenesis (GO:1902366) | 0.00003221 |
| oocyte anterior/posterior axis specification (GO:0007314) | 0.00003221 |
| regulation of somitogenesis (GO:0014807) | 0.00005170 |
| somite rostral/caudal axis specification (GO:0032525) | 0.0001202 |

**Figure 2.12: Pathway and biological process enrichment for gene module in 2.11** The enriched pathways and biological processes include the differentiation of several mesendoderm derivatives including cardiac, endoderm and mesoderm. Several terms related to somitogenesis are significantly enriched

**Figure 2.13: Example of module for genes that were upregulated late in the L and S lines only**

A. Hierarchical clustering of all the “late” genes that overlap in only the L and S. The colored bar at the bottom represents the separate modules detected by the dynamic tree-cutting algorithm. (Langfelder, 2008)
B. Heat map of the “black” group as detected above. Genes have a distinct expression pattern across the L and S lines; all the genes are upregulated at day 25 in the L line, versus only at day 30 in the S line. This group contains several well-known genes related to myogenic differentiation: MYOG, MOYD1, MYH8, PITX3, MYOT, MYH3, and CHRNA1. This group includes the lncRNA LINCMD1.


LS_only_late
Height
0
5
10
15
20
25
30
DS3
Color Key
0 2 4 6 8
Value
LS_only_late:black
CHRNA1
LMOD3
MYH3
TNFRSF13B
HSPB3
LINCMD1
AC020909.2
AC083902.2
ASPN
UNC45B
RAPSN
MYF5
AKNAD1
CASS4
HFE2
VGLL2
CAV3
MYOT
MYH7B
LINC00514
AC016717.2
ACADL
EDN3
PITX3
C4A
IL17B
SGCA
MYL1
MYH8
MYOD1
MYOG
RTL1
TMEM8C
KRT5
MEG3
D3T
D6T
D8T
D12T
D16T
D19T
D25T
D30T
D3S
D6S
D8S
D12S
D16S
D19S
D25S
D30S
D3L
D6L
D8L
D12L
D16L
D19L
D25L
D30L

| Wikipathways 2016 | p-value |
|---|---|
| Striated Muscle Contraction_Homo sapiens_WP383 | 0.00003021 |
| miRs in Muscle Cell Differentiation_Homo sapiens_WP2012 | 0.001682 |
| ID signaling pathway_Homo sapiens_WP53 | 0.02530 |
| Mitochondrial LC-Fatty Acid Beta-Oxidation_Homo sapiens_WP368 | 0.02687 |
| Hypertrophy Model_Homo sapiens_WP516 | 0.03153 |

| GO Biological Process 2017b | p-value |
|---|---|
| branchiomeric skeletal muscle development (GO:0014707) | 1.481e-9 |
| muscle filament sliding (GO:0030049) | 1.742e-8 |
| muscle organ development (GO:0007517) | 2.725e-8 |
| pharyngeal muscle development (GO:0043282) | 2.725e-8 |
| muscle cell fate determination (GO:0007521) | 1.297e-7 |
| skeletal muscle satellite cell differentiation (GO:0014816) | 2.072e-7 |

**Figure 2.14: Pathway and biological process enrichment for gene module in 2.12**
The enriched pathways and biological processes include terms that reflect myogenic differentiation and skeletal muscle contraction. A term for satellite cell differentiation is also included.

| Names | Category (encode) | Location in genome | # Transcripts and length | Overlap protein? |
|---|---|---|---|---|
| AC020909.2 | lincRNA | Ch19: 50,486,810-50,487,638 RevStrand | 1 transcript, 626 BP, No orthologues | Yes, EMC10 exon; |
| AC083902.2 | Protein Coding | Ch4: 99,636,529-99,654,648 RevStrand | 1 transcript, 10,086 bp 42 orthologues | No |
| LINC00514 | Retained intron, Processed transcript | Ch 16: 2,988,961-2,994,509 FwdStrand | 3 transcripts: 1)3385 bp; 2)2163)313 | Overlaps AC004233.2 |
| AC016717.2 | lincRNA | Ch2: 225,698,514-225,703,654 FwdStr | 1 transcript, 5141 bp | No |

**Figure 2.15: Unannotated transcripts that cluster with myogenic regulatory factors shown in Figure 2.13**

The four unannotated transcripts above cluster closely with numerous myogenic regulatory factors as detailed in Figure 2.13B. We have tabulated their predicted products according to encode, genomic location and other details about the transcripts.

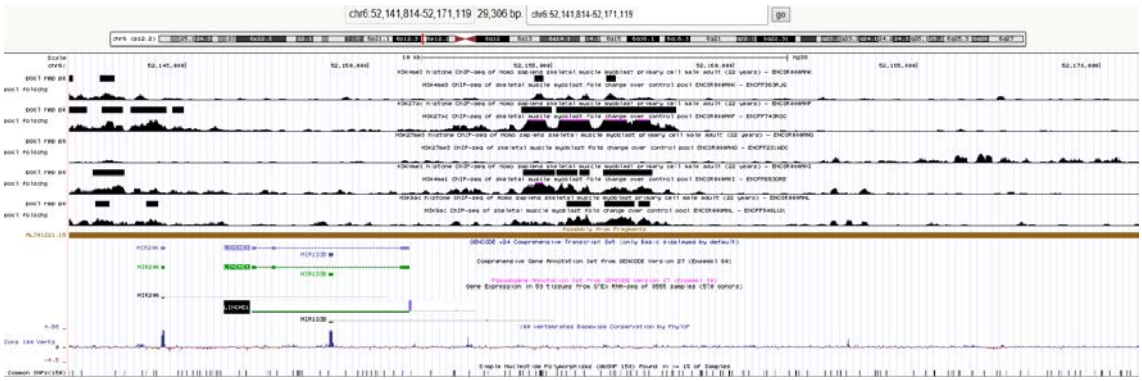


**Figure 2.16: Genome browser with integrated CHiP seq and tissue specific transcript expression data for LINCMD1**

We show this publicly available data for a long noncoding RNA known to play a role in promoting myogenic differentiation as a "positive control". The CHiP seq data shows enriched peaks in activating histone modifications upstream of the transcript of interest. Furthermore, the tissue specific gene expression shows that lincMD1 transcripts are present in skeletal muscle, but not other tissue types.

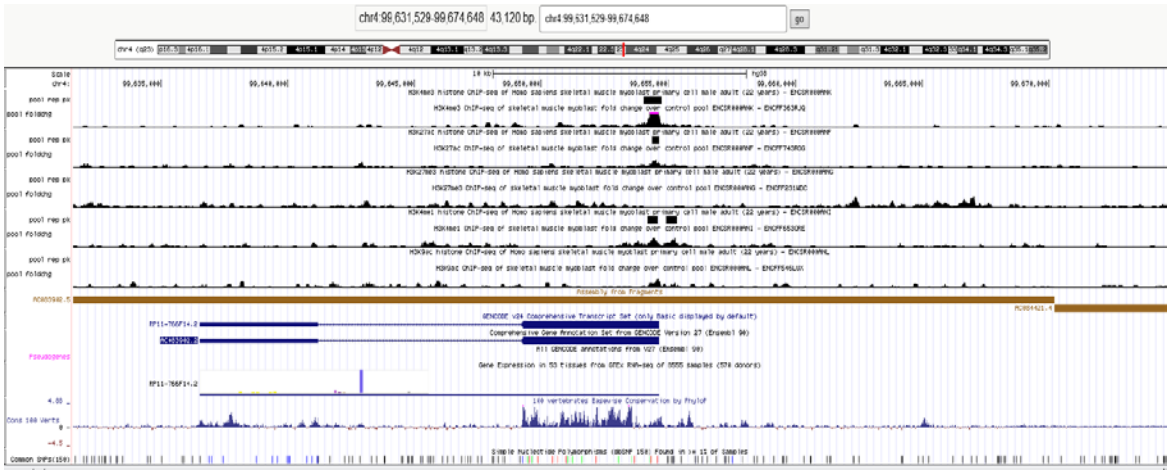



**Figure 2.17: Genome browser with integrated CHiP seq and tissue specific transcript expression data for ACO83902.2**

Similar to Figure 2.16, this is the publicly available CHiP seq and tissue specific transcript expression data for unannotated transcript ACO83902.2; encode classifies it as a putative protein coding transcript. The CHiP seq data shows enriched peaks in activating histone modifications, namely H3K4me3, upstream of the transcript of interest. Furthermore, the tissue specific gene expression shows that ACO83902.2 transcripts are present in skeletal muscle, but not other tissue types.

## References

Aguilar, C.A., Shcherbina, A., Ricke, D.O., Pop, R., Carrigan, C.T., Gifford, C.A., Urso, M.L., Kottke, M.A., Meissner, A., 2015. In vivo Monitoring of Transcriptional Dynamics After Lower-Limb Muscle Injury Enables Quantitative Classification of Healing. Scientific Reports 5, 13885.

Allegrucci, C., Young, L.E., 2007. Differences between human embryonic stem cell lines. Hum Reprod Update 13, 103–120.

Bilic Josipa, Belmonte Juan Carlos Izpisua, 2011. Concise Review: Induced Pluripotent Stem Cells Versus Embryonic Stem Cells: Close Enough or Yet Too Far Apart? STEM CELLS 30, 33–41.

Borchin, B., Chen, J., Barberi, T., 2013. Derivation and FACS-Mediated Purification of PAX3+/PAX7+ Skeletal Muscle Precursors from Human Pluripotent Stem Cells. Stem Cell Reports 1, 620–631.

Braun, T., Gautel, M., 2011. Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. Nature Reviews Molecular Cell Biology 12, 349–361.

Buchberger, A., Nomokonova, N., Arnold, H.-H., 2003. Myf5 expression in somites and limb buds of mouse embryos is controlled by two distinct distal enhancer activities. Development 130, 3297–3307.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., Bozzoni, I., 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147, 358–369.

Chal, J., Al Tanoury, Z., Hestin, M., Gobert, B., Aivio, S., Hick, A., Cherrier, T., Nesmith, A.P., Parker, K.K., Pourquié, O., 2016. Generation of human muscle fibers and satellite-like cells from human pluripotent stem cells in vitro. Nat Protoc 11, 1833–1850.

Chal, J., Oginuma, M., Tanoury, Z.A., Gobert, B., Sumara, O., Hick, A., Bousson, F., Zidouni, Y., Mursch, C., Moncuquet, P., Tassy, O., Vincent, S., Miyanari, A., Bera, A., Garnier, J.-M., Guevara, G., Hestin, M., Kennedy, L., Hayashi, S., Drayton, B., Cherrier, T., Gayraud-Morel, B., Gussoni, E., Relaix, F., Tajbakhsh, S., Pourquié, O., 2015. Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. Nature Biotechnology 33, 962–969.

Daubas, P., Tajbakhsh, S., Hadchouel, J., Primig, M., Buckingham, M., 2000. Myf5 is a novel early axonal marker in the mouse brain and is subjected to post-transcriptional regulation in neurons. Development 127, 319–331.

Goudenege, S., Lebel, C., Huot, N.B., Dufour, C., Fujii, I., Gekas, J., Rousseau, J., Tremblay, J.P., 2012. Myoblasts derived from normal hESCs and dystrophic hiPSCs efficiently fuse with existing muscle fibers following transplantation. Mol. Ther. 20, 2153–2167.

He, K., Wu, G., Li, W.-X., Guan, D., Lv, W., Gong, M., Ye, S., Lu, A., 2017. A transcriptomic study of myogenic differentiation under the overexpression of PPARγ by RNA-Seq. Scientific Reports 7, 15308.

Huang, Y., Wan, J., Guo, Y., Zhu, S., Wang, Y., Wang, L., Guo, Q., Lu, Y., Wang, Z., 2017. Transcriptome Analysis of Induced Pluripotent Stem Cell (iPSC)-derived Pancreatic β-like Cell Differentiation. Cell Transplant 26, 1380–1391.

Huggins, I.J., Bos, T., Gaylord, O., Jessen, C., Lonquich, B., Puranen, A., Richter, J., Rossdam, C., Brafman, D., Gaasterland, T., Willert, K., 2017. The WNT target SP5 negatively regulates WNT transcriptional programs in human pluripotent stem cells. Nature Communications 8, 1034.

Hu, S., Zhao, M.-T., Jahanbani, F., Shao, N.-Y., Lee, W.H., Chen, H., Snyder, M.P., Wu, J.C., 2016. Effects of cellular origin on differentiation of human induced pluripotent stem cell–derived endothelial cells. JCI Insight 1.

Hwang, Y., Suk, S., Lin, S., Tierney, M., Du, B., Seo, T., Mitchell, A., Sacco, A., Varghese, S., 2013. Directed In Vitro Myogenesis of Human Embryonic Stem Cells and Their In Vivo Engraftment. PLOS ONE 8, e72023.

Kabra, H., Hwang, Y., Lim, H.L., Kar, M., Arya, G., Varghese, S., 2015. Biomimetic Material-Assisted Delivery of Human Embryonic Stem Cell Derivatives for Enhanced In Vivo Survival and Engraftment. ACS Biomater Sci Eng 1, 7–12.

Kumar, N., Richter, J., Cutts, J., Bush, K.T., Trujillo, C., Nigam, S.K., Gaasterland, T., Brafman, D., Willert, K., 2015. Generation of an expandable intermediate mesoderm restricted progenitor cell line from human pluripotent stem cells. Elife 4.

Langfelder, P., Zhang, B., Horvath, S., 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24, 719–720.

Schoenwolf, G.C., Bleyl, S.B., Brauer, P.R., Francis-West, P.H. & Philippa H. (2015). *Larsen's human embryology* (5th ed.). New York; Edinburgh: Churchill Livingstone

Li, Y., Lin, B., Yang, L., 2015. Comparative Transcriptomic Analysis of Multiple Cardiovascular Fates from Embryonic Stem Cells Predicts Novel Regulators in Human Cardiogenesis. Sci Rep 5.

Li, Y., Wang, R., Qiao, N., Peng, G., Zhang, K., Tang, K., Han, J.-D.J., Jing, N., 2017. Transcriptome analysis reveals determinant stages controlling human embryonic stem cell commitment to neuronal cells. J. Biol. Chem. 292, 19590–19604.

Mesnard, D., Donnison, M., Fuerer, C., Pfeffer, P.L., Constam, D.B., 2011. The microenvironment patterns the pluripotent mouse epiblast through paracrine Furin and Pace4 proteolytic activities. Genes Dev. 25, 1871–1880.

Nichols, J., Smith, A., 2009. Naive and primed pluripotent states. Cell Stem Cell 4, 487–492.
Nishino, K., Toyoda, M., Yamazaki-Inoue, M., Fukawatase, Y., Chikazawa, E., Sakaguchi, H., Akutsu, H., Umezawa, A., 2011. DNA Methylation Dynamics in Human Induced Pluripotent Stem Cells over Time. PLoS Genet 7.

Pera, M.F., 2010. Defining pluripotency. Nature Methods 7, 885–887.

Rao, L., Qian, Y., Khodabukus, A., Ribar, T., Bursac, N., 2018. Engineering human pluripotent stem cells into a functional skeletal muscle tissue. Nature Communications 9, 126.

Rizzo, F., Ramirez, A., Compagnucci, C., Salani, S., Melzi, V., Bordoni, A., Fortunato, F., Niceforo, A., Bresolin, N., Comi, G.P., Bertini, E., Nizzardo, M., Corti, S., 2017. Genome-wide RNA-seq of iPSC-derived motor neurons indicates selective cytoskeletal perturbation in Brown–Vialetto disease that is partially rescued by riboflavin. Scientific Reports 7, 46271.

Shan, T., Liang, X., Bi, P., Zhang, P., Liu, W., Kuang, S., 2013. Distinct populations of adipogenic and myogenic Myf5-lineage progenitors in white adipose tissues. J Lipid Res 54, 2214–2224.

Tompkins, J.D., Jung, M., Chen, C.Y., Lin, Z., Ye, J., Godatha, S., Lizhar, E., Wu, X., Hsu, D., Couture, L.A., Riggs, A.D., 2016. Mapping Human Pluripotent-to-Cardiomyocyte Differentiation: Methylomes, Transcriptomes, and Exon DNA Methylation “Memories”., Mapping Human Pluripotent-to-Cardiomyocyte Differentiation: Methylomes, Transcriptomes, and Exon DNA Methylation “Memories.” EBioMedicine 4, 4, 74, 74–85.

Wang, Y., Zhao, C., Hou, Z., Yang, Y., Bi, Y., Wang, H., Zhang, Y., Gao, S., n.d. Unique molecular events during reprogramming of human somatic cells to induced pluripotent stem cells (iPSCs) at naïve state. eLife 7.

Xi, H., Fujiwara, W., Gonzalez, K., Jan, M., Liebscher, S., Handel, B.V., Schenke-Layland, K., Pyle, A.D., 2017. In Vivo Human Somitogenesis Guides Somite Development from hPSCs. Cell Reports 18, 1573–1585.

Young, C.S., Hicks, M.R., Ermolova, N.V., Nakano, H., Jan, M., Younesi, S., Karumbayaram, S., Kumagai-Cresse, C., Wang, D., Zack, J.A., Kohn, D.B., Nakano, A., Nelson, S.F., Miceli, M.C., Spencer, M.J., Pyle, A.D., 2016. A Single CRISPR-Cas9 Deletion Strategy that Targets the Majority of DMD Patients Restores Dystrophin Function in hiPSC-Derived Muscle Cells. Cell Stem Cell 18, 533–540.

# Chapter 3: Mechanisms of *in vitro* Myogenesis of Human Induced Pluripotent Stem Cells, part 2

## Introduction

In the previous chapter, we observed a divergence in the gene expression profiles of the L cell line from the S and T cell lines, mirroring the accelerated appearance of markers of myogenic specification in the L line starting around day 19 of differentiation. Through clustering analysis of the time points from all three cell lines, we noted that the gene expression profiles of the three cell lines actually started to diverge much earlier, around day 6. Therefore, we hypothesized that a difference in initial germ layer specification could lead to the observed temporal and extent of commitment shift in terminal myogenic specification. In particular, perhaps the L line that showed accelerated myogenesis had more efficient or accelerated early specification to the paraxial mesoderm lineage than the other two cell lines.

To test this, we first examined the expression of early germ layer specifiers and their downstream targets that mediate terminal specification. Expression of genes related to initial germ layer specification differed between the L line compared to the S and T lines. In particular, we observed expression of the gene EOMES and other early endoderm genes in the S and T lines, but not the L line. Conversely, we observe longer upregulation of genes related to mesoderm specification in the L line, including brachyury. Using the longitudinal data, we follow these differences in initial germ layer specification farther in time to see that transcription factors that regulate the specification of lineages that derive

from either endoderm or mesoderm are upregulated later in time in the corresponding cell lines.

We took a systems biology perspective to better understand the reciprocal interaction of multiple factors, including chromatin structure modification, the activity of major pathways, and the core transcriptional regulators that integrate these signals to effect control of gene expression and cellular identity. First, we used the clustering strategy described in the previous chapter to identify groups of genes related to pluripotency and early lineage specification and looked for the enriched transcription factors corresponding to each group using the ChEA database. We paid particular attention to the cell-line dependent expression of the regulators of chromatin structure and histone state, and the extra- and intra-cellular components of several major signaling pathways. Multiple factors that regulate chromatin structure, including chromatin modifying complexes and long noncoding RNAs, regulate cellular identity in an extremely context-specific manner by controlling transcription factor binding and subsequent gene expression. Our longitudinal dataset that captures the more efficient myogenic specification of one hiPSC line compared to two others helps contextualize the observed cell-line dependent variability in the expression of these regulators of chromatin structure.

Next, we compared the cell-line dependent expression of key components of major signaling pathways. Modulation of signaling pathways such as Notch, Wnt, BMP, and Activin/Nodal by small molecules and growth factors are the foundation for numerous differentiation protocols for hPSCs, including the one we used. During embryological development, ligand gradients guide morphogenesis. We observed differential, cell-line

varying expression of members of several major pathways that corresponds well with known patterns of pathways activity that lead to endoderm or mesoderm specification. Finally, we considered the core transcriptional machinery that lies downstream from and integrates signals from several signaling pathways to affect gene transcription. In particular, we consider the transcriptional repressors and co-factors of beta catenin, since several differentiation protocols for mesoderm and endoderm specification, including the one we used, rely on wnt activation via increased beta catenin translocation to the nucleus to achieve initial germ layer specification. We compared the longitudinal cell-line dependent expression of several of these cofactors, with an eye towards genetic perturbation of several of these genes that could either enhance or blunt paraxial mesoderm commitment and downstream myogenic differentiation.

## Results

### Cell line-dependent initial germ layer commitment and subsequent expression of downstream lineage specific transcription factors

We observed divergent gene expression profiles between the L line compared to the S and T lines early in differentiation. As a result, we first investigated the expression patterns of genes that regulate pluripotency and early lineage specification. (Figure 3.7) Among the genes that are known to be active in pluripotency, POU5F1 and SOX2 had similar patterns of downregulation in all three lines, while KLF4 and NANOG did not. In particular, NANOG was strongly downregulated in the L line, but not in the T and S lines. We also observed a similar cell-line dependent expression pattern in key genes that regulate

mesendoderm specification. Expression of genes related to initial germ layer specification differ between L line and S and T lines, in particular we observe expression of the gene EOMES in the S and T lines, but not the L line. Other transcription factors more associated with endoderm than mesoderm specification such as FOXA2 and GATA4 were upregulated early in differentiation in the S and T lines, but had no to little upregulation in the L lines. By later differentiation- days 19, 25, and 30- the S and T lines express SOX17, a marker of definitive endoderm. In addition, they express HFN4A, hepatocyte nuclear factor 4 alpha, a transcription factor that may play a role in liver development, a derivative of definitive endoderm. By day 30, the S and T lines express SERPINA1 that encodes alpha 1 antitrypsin, a protease inhibitor produced in the liver. Genes related to cardiac development are also generally more upregulated in the S and T lines than the L line. The cardiac specification transcription factor NKX2-5 is up regulated in the S line, and HAND1 is most upregulated in the S and T lines, although it is also upregulated in the L line.

Conversely, we observe longer upregulation of genes related to mesoderm specification in the L line, including brachyury and MSGN1. PAX3, a transcription that plays a role in somitic dermomytome development, is only transiently upregulated in the S and T lines, while it is strongly upregulated throughout differentiation in the L line. There is a similar expression pattern for MEOX1, another gene involved in somite development, although it is more associated with sclerotome development than dermomyotome. In addition to regulators of paraxial mesoderm development, the L line also had increased upregulation of genes such as related to intermediate mesoderm (OSR1, WT1), which gives rise to the nephric system. Finally, the L line had the most upregulation of myogenic

regulatory factors, including MyoG starting around day 19. The S line had only relatively moderate MyoG upregulation by day 30, while the T line did not have MyoG upregulation in 30 days of differentiation.

Identification of gene modules that regulate pluripotency and initial germ layer specification

Since it appears that lineage commitment diverged early in differentiation, it would be important to understand the transcription factors that regulate the transition from pluripotency to initial germ player. We identified groups of genes that clustered with transcriptional regulators of pluripotency and initial germ layer specification using the hierarchical clustering with dynamic tree cut strategy described in the previous chapter. (Figure 3.1) We used SOX2, POU5F1, PRDM14, NANOG, and KLF4 as key regulators of pluripotency and selected the genes that clustered with each of these, as they had similar patterns of expression across the three cell lines. Although MYC is a Yamanaka factor used in the reprogramming of somatic cells to hiPSCs, we did not use MYC as a key regulator of pluripotency, as it plays a greater role in self-renewal than reprogramming per se. As such, MYC is active in more contexts outside pluripotency compared to the other Yamanaka factors, NANOG, or PRDM14. The SOX2-associated gene group has a similar expression pattern across the three cell lines- the genes are downregulated early in differentiation. Notable genes that cluster with SOX2 are SOX21, HDAC9, and DEPTOR. In the POU5F1-related group, the genes are downregulated in all three cell lines, but there

is relatively increased downregulation in the L line. ZFP42 (Rex1) clusters with POU5F1. Interestingly, TERF1 a regulator of telomere length also clusters with POU5F1 and ZFP42.

The expression pattern across cell lines diverges in the PRDM14-, NANOG-, and KLF4- associated groups. Several genes associated with PRDM14, including TDGF1 and ZSCAN10 are extensively studied in the context of pluripotency. Another telomere regulator, TERT, is also clustered in the PRDM14-associated group. Compared to the POU5F1-associated group, these genes are strongly downregulated in the L line and partially downregulated in the S and T lines. Similarly, the genes associated with KLF4 are more strongly and consistently downregulated in the L line, and intermittently to not downregulated in the T and S lines, respectively. This trend is the most marked in the NANOG-associated gene module. The NANOG cluster contains LINCPRESS1 and 2, which are long noncoding RNAs that play a role in the regulation of pluripotency. Notably, the endogenous retrovirus associated gene ERVH-1 clusters with NANOG; endogenous retrovirus-related gene expression has been described as potential species-specific regulators of pluripotency, so it is interesting that this gene clusters closely with NANOG.

We used T (brachyury), EOMES, HOXA2, HOXB5, and HOXD10 as key lineage-specification genes. Brachyury, a pan-primitive streak marker, is upregulated in progenitors of all mesodermal and endodermal subtypes. Several important transcription factors involved in mesoderm specification are in this cluster, including MIXL1, TBX6, and CDX2. FGF17 and WNT8A, components of two major signaling pathways are likewise in this cluster. Genes in the T-associated cluster are upregulated early across all three cell lines, but have longer temporal expression in the L line compared to the other

two lines. EOMES is also an early marker associated with mesendodermal progenitors. The EOMES-associated genes are generally upregulated in all three cell lines. Interestingly, EOMES and the genes that cluster most closely with it are strongly upregulated early in the differentiation of the S and T lines, but not at all in the L line. It is of note that one of these genes is LHX1, a key transcription factor in endoderm specification.

The HOX gene family has a characteristic rostral to caudal expression pattern in the developing embryo. We have selected HOXA2 as a representative of rostral fates, HOXB5 as a representative of middle fates, and HOXD10 for the most caudal. The gene group associated with HOXA2 is upregulated with similar pattern across all three cell lines; this group also includes SNAI2 (Slug), a regulator of endothelial-mesenchymal transition (EMT) that is required for ingression through the primitive streak. The HOXB5 associated group is also upregulated across all three cell lines, but they are more upregulated in the L line than in the other two lines. This group includes several more rostral/middle HOX genes, as well as MEOX1, EDAR, and DKK1, which have been associated with various aspects of mesendoderm specification. Finally, the HOXD10-associated group is upregulated in the L line, particularly in the latter half of differentiation, but weak to no upregulation in the T and S lines. The varying expression patterns across the three cell lines of gene modules associated with more anterior or posterior embryonic position may reflect differing initial germ layer specification across the three cell lines.

The functional enrichment of both groups of genes confirms their roles in regulating either pluripotency or differentiation. (Figure 3.2) The pluripotency related group was

enriched for terms including Wnt Signaling and pluripotency, preimplantation embryo, mesoderm commitment, and endoderm differentiation. The differentiation group was also enriched for terms such as neural crest, endoderm, and cardiac differentiation, as well as mesodermal commitment and wnt signaling.

Core components of major chromatin and histone modification complexes are enriched upstream of key groups of genes involved in the regulation of pluripotency and lineage specification

We performed transcription factor (TF) enrichment analysis with Enrichr, using the Chea database for both pluripotency related genes and differentiation associated genes separately and compared the enriched TFs of both groups. (Figure 3.3) Enriched TFs for genes regulating pluripotency include SMAD3, RUNX, and PAX3, which reflects their known role in meditating exit from pluripotency. Similarly, TFs enriched in the differentiation-associated set include CTNNB1, CDX2, and WT1. These TFs also reflect activity of the canonical WNT/B-catenin and TGFB-family/SMAD pathways, which is consistent with the small molecules CHIR and LDN that are added in the first step of the differentiation protocol to induce specification to paraxial mesoderm. However, as these pathways play context-specific roles in differentiation, their activity does not necessarily mean paraxial specification. Enriched TFs from both groups contained key core components of major chromatin remodeling complexes that are known to play a role in differentiation, including PRC2, PRC1, NuRD. Interestingly, although the core components are enriched in both groups several ancillary components of these complexes

are enriched in either the pluripotency-related group (EP300, REST) or in the differentiation-associated group (BMI1, CHD7, PHC1). This may reflect the context specific compositions of major chromatin modifying complexes, which may have implications for germ layer commitment.

Cell line dependent variations in the temporal expression of the components of key histone and chromatin modifying complexes

To further investigate whether there may be cell line dependent differences in the composition of major chromatin complexes, we investigated the gene expression patterns of the components of the complexes PRC1, PRC2,(Figure 3.4) SWI/SNF and NuRD (Figure 3.5). We show the core components as well as the changeable components of these complexes. In most cases, with the exception of PRC1which has several canonical and non-canonical forms described in the literature, the core components of the other are not significantly differentially expressed. However, several other components that associate with the core complex and may play a role in targeting the complex to context specific regions of the genome have expression patterns that roughly correspond with the overall pattern of gene expression between the cell lines- the L line is different from the S and T lines. The components of PRC2 are not significantly DE, with the exception of JARID2, which is much more downregulated in the L line compared to the other two lines. This is consistent with reports in the literature, which describe the role of JARID2 in PRC2 binding in pluripotent cells. Similarly, several components of PRC1 are DE between cell lines. The core of the SWI/SNF complex is comprised of SMARCA4 (Brg1) or SMARCA2 (Brm).

SMARCA4 is widely studied in the context of cancer, with some study in the context of pluripotency, while SMARCA2 is less widely studied, with a recent report comparing the potential role of SMARCA2 versus SMARCA4 in the metastatic potential of lung cancers. In this context, the composition of the SNI/SNF complex core is thought to affect the grade of cancer (ie stemness), so it is interesting to note that SMARCA2 is downregulated early in the L line, but not the other two. Finally, the TRIM28-associated NuRD complex has a similar pattern of differential expression of some components between the cell lines. In particular, TRIM28 is well known to play a role in promoting pluripotency; it is most heavily downregulated in the L line, which had the most robust myogenesis. Furthermore, many zinc finger transcription factors that interact with NuRD/TRIM28 exhibit the same characteristic expression pattern between the cell lines as do the major known genes that regulate pluripotency. This is potentially interesting as the function of only few of this family of transcription factors has been studied.

Annotated long non-coding RNAs and unannotated transcripts cluster with regulators of pluripotency and lineage specification

Since clustering genes together across cell lines is an effective way to group together genes that play a role in the same biological context, we took used this strategy to contextualize unannotated transcripts, including several long non-coding RNAs. Figure 3.6A is an example of known and unknown lncRNAs that clustered together with protein coding genes known to function in the context of pluripotency and differentiation. In particularly, the HOX gene cluster is well-known for being regulated by numerous

lncRNAs, including HOTAIR which interacts with PRC2. Several unannotated transcripts cluster very close to these genes, perhaps indicating their role in regulating HOX or related-genes. Various known and unknown noncoding RNAs cluster closely with NANOG, in particular, LINCPRESS1 and 2 as well as LINC01108 (ESC-associated 1). The lncRNA PANCR (PITX6-associated) clusters closely with LHX1, a well-known regulator of definitive endoderm specification. Although PITX6 is known to function in the context of endoderm differentiation, downstream of Activin/Nodal signaling, PANCR is not widely studied; the gene expression pattern of PANCR across cell lines may indicate its role in regulating definitive-endoderm specification. We also note that genes related with endoderm specification are strongly upregulated in the S and T cell lines, but not in the L line. Finally, numerous Histone 1H components are DE between the cell lines. Histone 1H components play a role in packaging chromatin around the core histone octamer and may play a role in determining the accessibility of chromatin. Several of these components are more downregulated in the T line than the other two lines, which is interesting given that the T line had the least myogenic differentiation in 30 days. (Figure 3.6B) Taken together, these cellular components- chromatin modifying complexes, lncRNAs, and histone H1 components- play important roles in determining chromatin structure, which has implications for gene expression patterns that lead to cellular identity.

Cell line dependent expression of the components of major signaling pathways correlate with germ layer specification

Relative activity of the major signaling pathways such as Wnt, Notch, BMP, Activin/Nodal, and Hedgehog plays an important role in determining cell fate. (Figure 3.8) Although RNA seq data does not shed light on the phosphorylation status of the intracellular signaling molecules and kinases, we can assess possible pathway activity by the up or down regulation of components of the pathways, especially extracellular ligands. The canonical Wnt pathway is an important pathway in paraxial mesoderm specification. The extracellular Wnt ligands Wnt3A, Wnt5B, and Wnt8A are initially upregulated in all three cell lines, consistent with Wnt activation by CHIR. However, in the L line, these genes have are upregulated longer and more intensely than in the S and T lines. In addition, the Frizzled receptor FZD10 is more highly upregulated in the L line than in the S and T lines.

There are notable differences in the expression of ligands of the TGFB family in the L line compared to the S and T lines. Nodal and Cerberus1 are up regulated by day 3 in the S and T lines; by contrast, Nodal is downregulated in the L line and Cer1 is not DE at day 3. Similarly, Lefty1 and activin A (INHBA) are more strongly down regulated in the L line and either weakly or not at all downregulated in the S and T lines. BMP4 is associated with cardiac and endoderm differentiation, and is more strongly upregulated in the S lie compared to the other two. LDN, an inhibitory small molecule that is specific for BMP Type1 receptors, is added to the differentiation for the first eight days of differentiation. However, BMP4 may have crosstalk with other types of receptors, and when LDN-mediated inhibition is removed after day 8, BMP4 would be able to signal through BMP Type1 receptors as well. AMHR2 (Anti-Mullerian hormone receptor) is a

TGFB superfamily receptor that is upregulated in the T and S lines only. Although AMHR2 is best known in its role in sex determination, a recent study has studied AMH/AMHR2 in the context of EMT in lung cancer. By contrast, the other TGFB family receptors, including BMP receptors, were not significantly differentially expressed. The co-receptor TDGF1 (Cripto) is known to play a role in maintaining pluripotency, and it is most strongly downregulated in the L line compared to the S and T lines.

There are similar patterns of expression in the ligands of the Notch signaling pathway. DLL1 and DLL3 are upregulated longer and more intensely in the L line than the S and T lines, in a pattern similar to the expression of Wnt3A/5B/8A, as well as that of FGF8/17/18. Interestingly, Notch1, is upregulated from D3 to D16 in the L line but only upregulated on D6 in the S and T lines. The Notch signaling pathway has been implicated in several developmental processes, especially neurogenesis.

Perturbation of transcription factors that interact with beta-catenin in the first three days of differentiation affects EOMES expression

B-catenin mediated transcription (by way of GSK3B inhibition by CHIR) - in a concentration-dependent, time-dependent manner that is subject to modulation by other signaling pathways- is the foundation of in vitro differentiation to several endoderm- and mesoderm-derived lineages. The formation of complexes with transcriptional repressors/activators and transcription factors that are downstream of other signaling pathways, such as SMADs serves as a way to allow for crosstalk and integration between signaling pathways. Additionally, B-catenin and its binding partners can interact with

chromatin modifying complexes further to regulate gene expression. (Figure 3.9A) Given its central role in lineage specification, we therefore asked whether there was cell-line dependent expression of the major B-catenin nuclear binding partners. We noted that several of these known interactors had distinct cell line dependent expression profiles that mirrored the expression patterns of other key lineage specification or pluripotency-related genes. (Figure 3.9B) HES7, that is downstream of the Notch pathway and is known to respond to GSK3B inhibition, has an expression pattern similar to brachyury. RBPJ, another transcriptional effector of Notch signaling, is downregulated in the L line, but not the other two lines. PYGO1 is upregulated in the L line, but either weakly of or not upregulated in the S and T lines. Members of the ZIC family of transcription factors, as well as several TLEs are differentially up and down regulated, respectively, in a cell line-dependent manner. Finally, although many of the TCF family transcription factors are not differentially expressed, LEF1 is more strongly upregulated in the L line compared to the other two lines. Taken together, this gives the impression of significant differences between the cell lines in the transcriptional network relating to beta-catenin and its cofactors.

To test whether perturbation of this network might lead to differences in lineage specification, we carried out an siRNA knockdown of 25 genes in the T line during the induction of differentiation. The genes were selected for their differential expression pattern between the cell lines early in differentiation, whether they are known to participate in the regulation of pluripotency or whether they interact with beta catenin to modulate its transcriptional activity. At day 3, we determined whether the gene knockdown condition led to more mesoderm or more endoderm specification by image quantification of

immunofluorescent staining for Brachyury and Eomes. (Figure 3.10) All the conditions had high brachyury expression, consistent with treatment with CHIR, and initial mesendoderm commitment. However, knockdown of the genes TLE6 and ZIC3 led to decreased Eomes expression compared to the negative control of scrambled siRNA. By contrast, Lef1 knockdown led to Eomes expression that was comparable to, or slightly higher than control. Taken together, we demonstrate that genetic perturbation of transcription factors that regulate beta catenin-mediated transcription can bias differentiation towards either endoderm or mesoderm.

## Discussion

In this chapter, we provide a mechanistic explanation of the broad trends in the data that we observed in the previous chapter using a systems biology approach. The temporal nature of the data was key to establishing a causal relationship between the relatively more efficient myogenesis in one cell line, as characterized by immunofluorescent staining, and the divergence in initial germ layer specification, as noted in the cell line dependent expression of key transcription factors that regulate lineage commitment. In particular, the cell line with robust myogenesis had greater commitment to a paraxial mesoderm fate while the other cell lines had fates that were more endodermal. These initial differences in lineage specification led to increasingly divergent progeny as the cells matured to terminal lineages that derived from either germ layer. (Figure 3.7)

A major difference between the cell lines was the difference in the downregulation of the key genes that regulate pluripotency. In particular, NANOG was heavily

downregulated in the L line, but much less so in the other cell lines. (Figure 3.1) This indicates that a subpopulation of cells may have remained as early progenitors instead of specifying to a terminally differentiated lineage. Furthermore, uneven downregulation of the regulators of pluripotency can bias differentiation. Nanog plays a role in specifying progenitors to an anterior primitive streak fate, while CDX2 plays a mutually exclusive role to specify a posterior primitive streak fate; this has downstream effects on the mesoderm subtype that arises. (Mandjan, 2014). In our experiment, the cells with higher Nanog expression preferentially went to an endoderm fate, which is derived from the anterior primitive streak. Interestingly, Nanog expression has been documented in adult fibroblast and smooth muscle cells in vitro (Ambady, 2010), though its function is unclear. Similarly, Oct4 is expressed in terminally differentiated human peripheral blood monocytes (Zangrossi, 2007), but this may be due to different isoforms of Oct4 (Kotoula, 2008). Although Sox2 is a Yamanaka factor and plays a key role together with Nanog and Oct4 in regulating the core pluripotency network, Sox2 also plays a role in directing neuroectoderm specification, at the expense of primitive streak differentiation (Wang, 2012; Li, 2017). These examples illustrate that the relative levels and interactions between the regulators of pluripotency bias differentiation.

In addition, specific subpopulations could effect the lineage commitment of neighboring cells through paracrine signaling and cell-cell contact. Paracrine effects of secreted Cer1 and Lefty1 were shown to bias hESC embryoid body differentiation towards endoderm, as opposed to mesoderm, in a cell density dependent manner (Kempf, 2016). Notch signaling via cell-cell contact between neural crest progenitors and somitic

dermomyotome guides embryonic skeletal muscle development (Rios, 2011); it has been hypothesized that a similar effect occurs during in vitro myogenic differentiation of hPSCs (Borchin, 2013). Similarly, pluripotent cells secrete Wnt ligands that help maintain pluripotency and self-renewal (Mills, 2017). Thus, a subpopulation of cells that remains in an early progenitor state may affect the lineage commitment of surrounding cells.

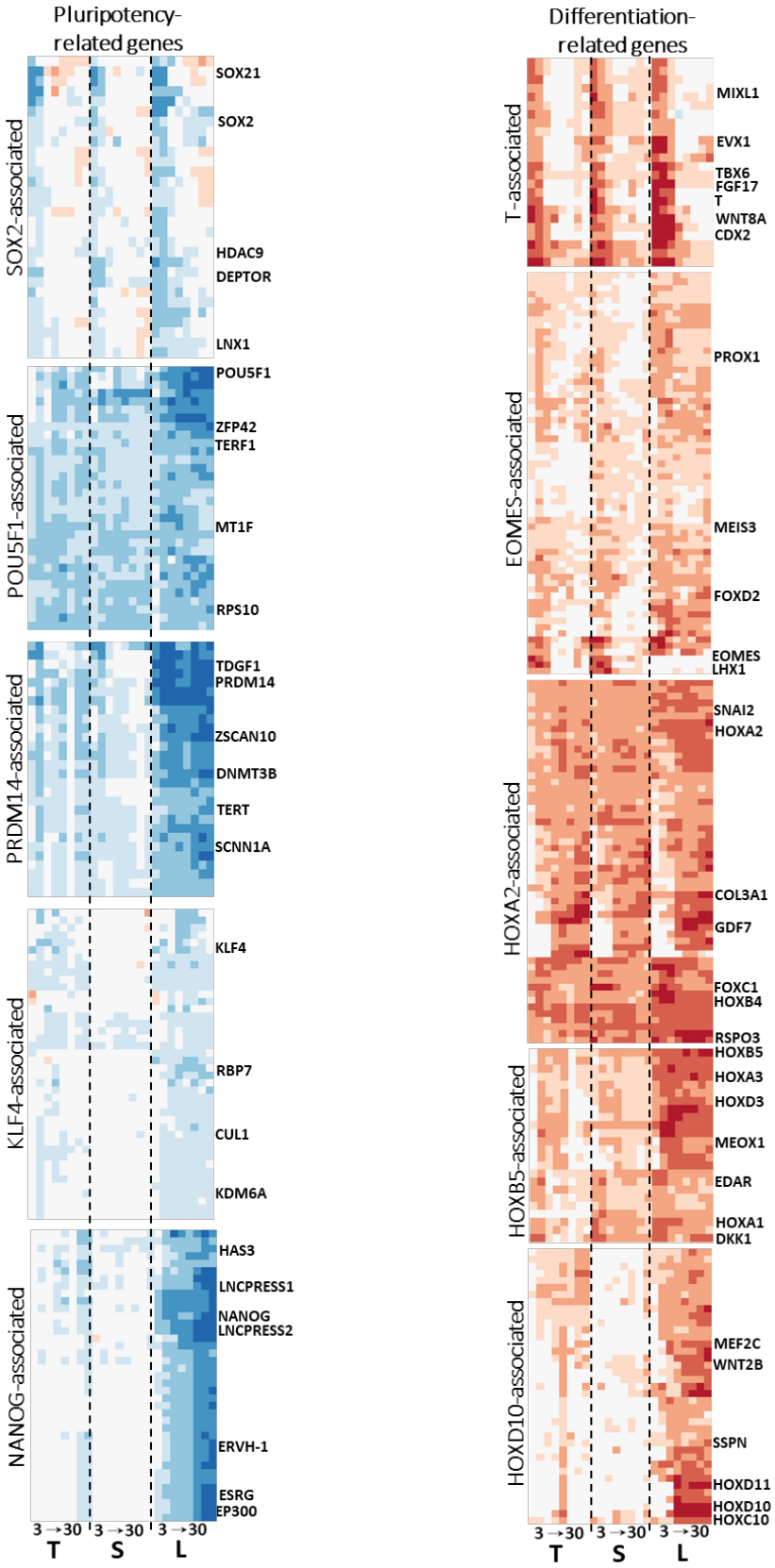
The pattern of HOX-family gene expression also supports the more posterior primitive streak/paraxial mesoderm fate in the line that had better myogenesis. The HOX genes are unique because their sequential genomic location mirrors their rostral to caudal expression in the developing embryo: The genes are numbered 1-13, with expression of the lower numbered HOX genes rostral, and limbs are caudal regions have higher numbers. (Goodman, 2013) In Figure 3.1, we selected three HOX genes: HOXA2, HOXB5, and HOXD10 to represent rostral, middle, and caudal expression patterns. We saw that the genes in the HOXA2-associated cluster were upregulated roughly equally in all three cell lines. The HOXB5 cluster genes were upregulated in all three lines, but had stronger upregulation in the L line. By contrast, the HOXD10-associated genes were strongly upregulated in the L line, but either weakly or not at all in the S and T lines. This pattern of expression suggests that the S and T lines have more anterior primitive streak origin, versus the L line, which expresses the posterior HOX genes. Paraxial and intermediate mesoderm are derived from the posterior streak; the L line expresses WT1 (kidney transcription factor, derived from intermediate mesoderm) as well as the myogenic markers.

In agreement with the literature, our analysis of transcription factor enrichment for the groups of genes related to pluripotency and differentiation heavily implicate regulators of the epigenome. However, relatively little is known about the details of how chromatin modifying complexes are targeted to specific segments of the genome in a context dependent way. Several studies document context and tissue specific compositions that may play a role in targeting (Zuqin, 2003). We examined the expression pattern across cell lines of the components of several major complexes, including the core enzymatic or ATP-ase components, as well as ancillary proteins, and found several components that were differentially expressed in the L line compared to the other two lines. Finally, we investigated the role of B-catenin cofactors in influencing early lineage specification with genetic perturbation.

## Figures

**Figure 3.1: Cell-line dependent expression profiles of genes modules related to pluripotency and differentiation**

T, S, and L hiPSC lines have variable up or down regulation for gene modules associated with key genes related to pluripotency and differentiation. SOX2, POU5F1, PRDM14, KLF4, and NANOG were selected for their key governing roles in the pluripotency gene regulatory networks. Likewise, T (brachyury) and EOMES were selected since they are early markers for mesendoderm specification. HOXA2, HOXB5, and HOXD10 were selected as representatives of transcriptional regulators of rostral to caudal body plan.


Pluripotency-related genes
SOX2-associated
SOX21
SOX2
HDAC9
DEPTOR
LNX1
POU5F1-associated
POU5F1
ZFP42
TERF1
MT1F
RPS10
PRDM14-associated
TDGF1
PRDM14
ZSCAN10
DNMT3B
TERT
SCNN1A
KLF4-associated
KLF4
RBP7
CUL1
KDM6A
NANOG-associated
HAS3
LNCPRESS1
NANOG
LNCPRESS2
ERVH-1
ESRG
EP300
3→30 3→30 3→30
T S L
Differentiation-related genes
T-associated
MIXL1
EVX1
TBX6
FGF17
T
WNT8A
CDX2
EOMES-associated
PROX1
MEIS3
FOXD2
EOMES
LHX1
HOXA2-associated
SNAI2
HOXA2
COL3A1
GDF7
FOXC1
HOXB4
RSPO3
HOXB5-associated
HOXB5
HOXA3
HOXD3
MEOX1
EDAR
HOXA1
DKK1
HOXD10-associated
MEF2C
WNT2B
SSPN
HOXD11
HOXD10
HOXC10
3→30 3→30 3→30
T S L

**A**

| Pluripotency-related enrichment terms | P-value |
|---|---|
| Wnt Signaling Pathway and Pluripotency_Homo sapiens_WP399 | 8.069e-4 |
| Preimplantation Embryo_Homo sapiens_WP3527 | 1.789e-3 |
| SIDS Susceptibility Pathways_Homo sapiens_WP706 | 2.689e-3 |
| Mesodermal Commitment Pathway_Homo sapiens_WP2857 | 5.389e-3 |
| Cardiac Progenitor Differentiation_Homo sapiens_WP2406 | 7.793e-3 |
| Notch Signaling Pathway_Homo sapiens_WP61 | 1.339e-2 |
| Endoderm Differentiation_Homo sapiens_WP2853 | 2.106e-2 |
| Cori Cycle_Homo sapiens_WP1946 | 2.285e-2 |
| Vitamin A and Carotenoid Metabolism_Homo sapiens_WP716 | 2.562e-2 |
| Imatinib Resistance in Chronic Myeloid Leukemia_Homo sapiens_WP2946 | 3.775e-2 |

**B**

| Differentiation-related enrichment terms | P-value |
|---|---|
| Neural Crest Differentiation_Homo sapiens_WP2064 | 1.202e-9 |
| Endoderm Differentiation_Homo sapiens_WP2853 | 5.160e-5 |
| Cardiac Progenitor Differentiation_Homo sapiens_WP2406 | 8.208e-5 |
| Senescence and Autophagy in Cancer_Homo sapiens_WP615 | 5.961e-4 |
| Endochondral Ossification_Homo sapiens_WP474 | 1.806e-3 |
| Mesodermal Commitment Pathway_Homo sapiens_WP2857 | 3.835e-3 |
| Dopaminergic Neurogenesis_Homo sapiens_WP2855 | 7.778e-3 |
| Ectoderm Differentiation_Homo sapiens_WP2858 | 1.217e-2 |
| Wnt Signaling Pathway_Homo sapiens_WP428 | 1.329e-2 |
| Primary Focal Segmental Glomerulosclerosis FSGS_Homo sapiens_WP2572 | 1.693e-2 |

**Figure 3.2: Pathways enrichment terms for pluripotency- and differentiation-related gene modules**

A. Enriched terms related to pluripotency include preimplantation embryo and Wnt signaling, in addition to differentiation to several lineages.
B. Enriched terms for differentiation-related genes contain differentiation to various lineages. Derivatives of all three germ layers are present.

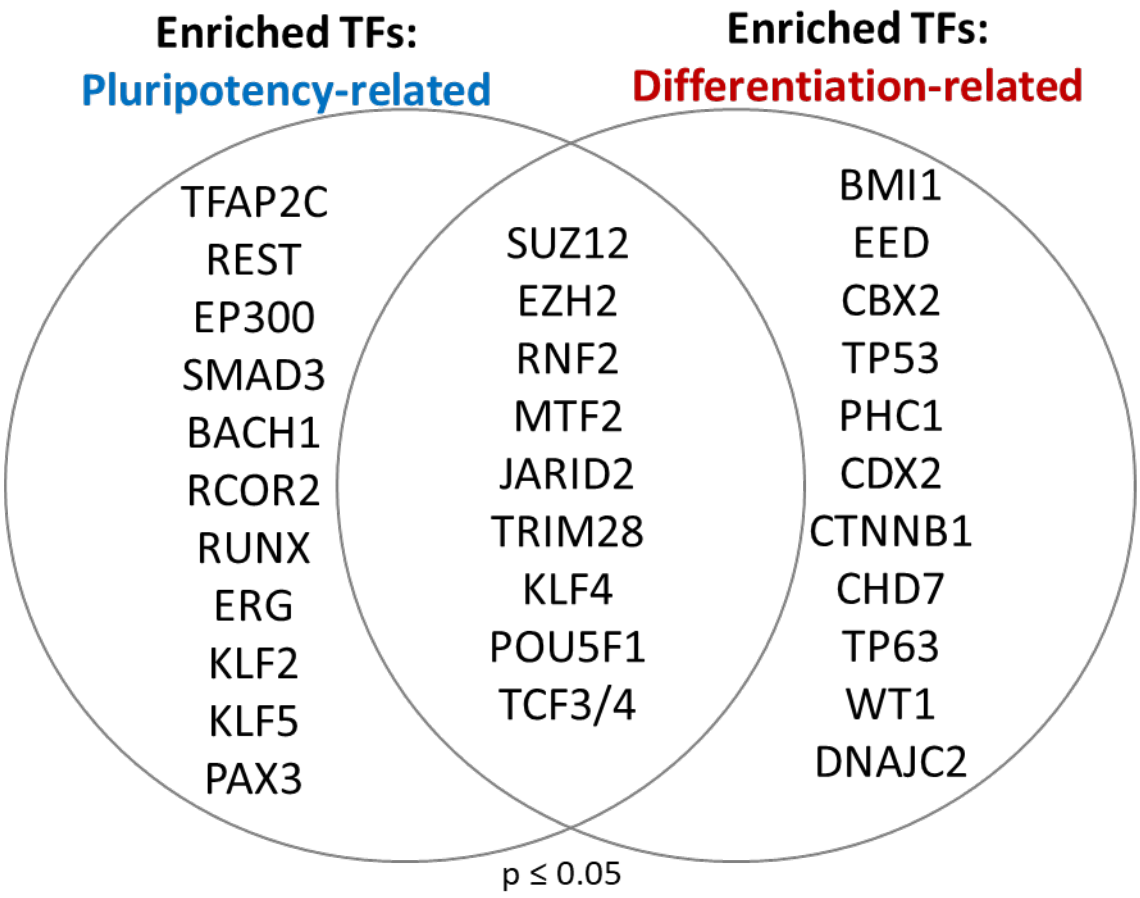



**Figure 3.3: Enriched transcription factors for pluripotency- and differentiation-related gene modules**

Enriched transcription factors for pluripotency- and differentiation related gene modules show that just under half of each group's enriched TFs overlap. Many of the enriched TFs are components of chromatin modifying complexes, including PRC2 core components (EED, SUZ12, EZH2), PRC2-associated JARID2 and PRC1 (RNF2, BMI1, CBX2, PHC1). In addition, CTNNB1 (B-catenin) is enriched for the differentiation-related set (at a p-value threshold of less than 0.05), but cofactors for B-catenin TCF3/4 are enriched for both gene modules.

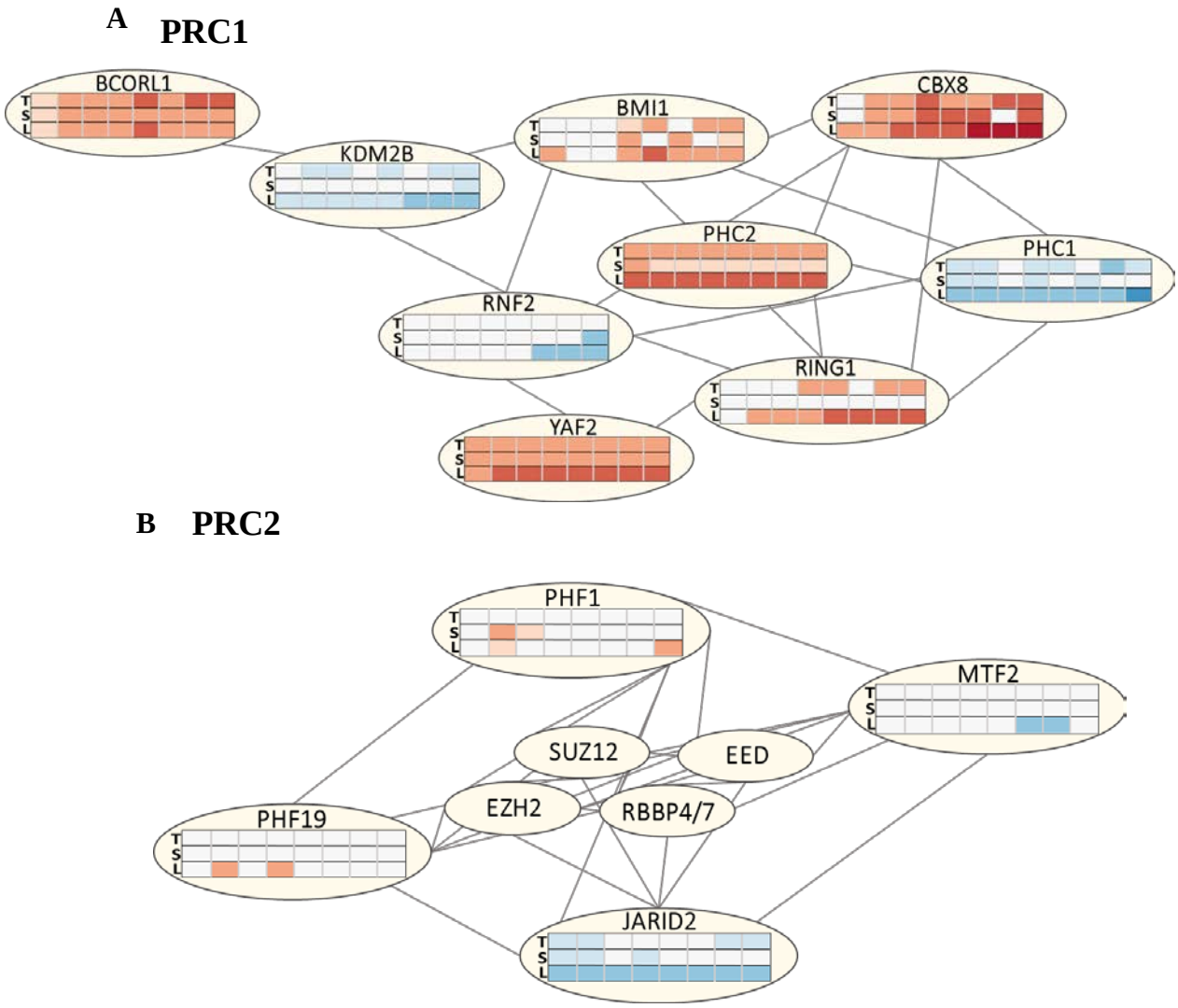



**Figure 3.4: Cell line-dependent gene expression profiles of components of Polycomb Repressor Complexes**

A. Polycomb Repressor Complex 1 (PRC1) has several components that exhibit cell line dependent expression patterns, including its core methyltransferases RNF2 and RING1. Components of canonical and noncanonical PRC1 are shown together.

B. Polycomb Repressor Complex 2 (PRC2) core components (EZH2, SUZ12, RBBP4/&, EED) were not found to be differentially expressed in any cell line. However, JARID2 is much more downregulated across time in the L line than the other two lines.

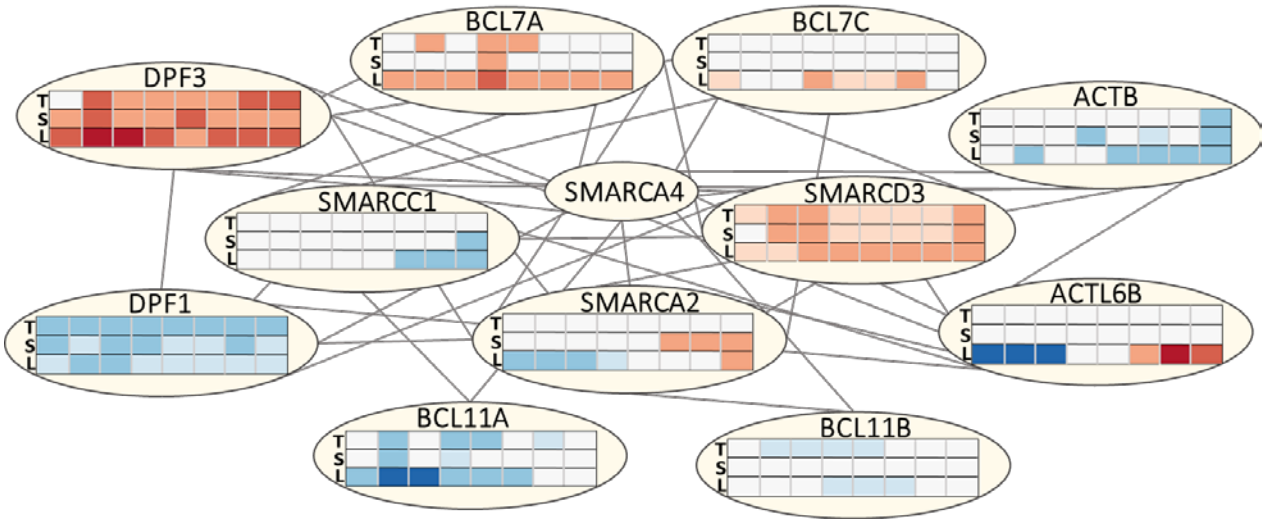
Lines between nodes are known protein-protein interactions from Stringdb.

**Figure 3.5: Cell line-dependent gene expression profiles of components of ATP-dependent chromatin remodelers**

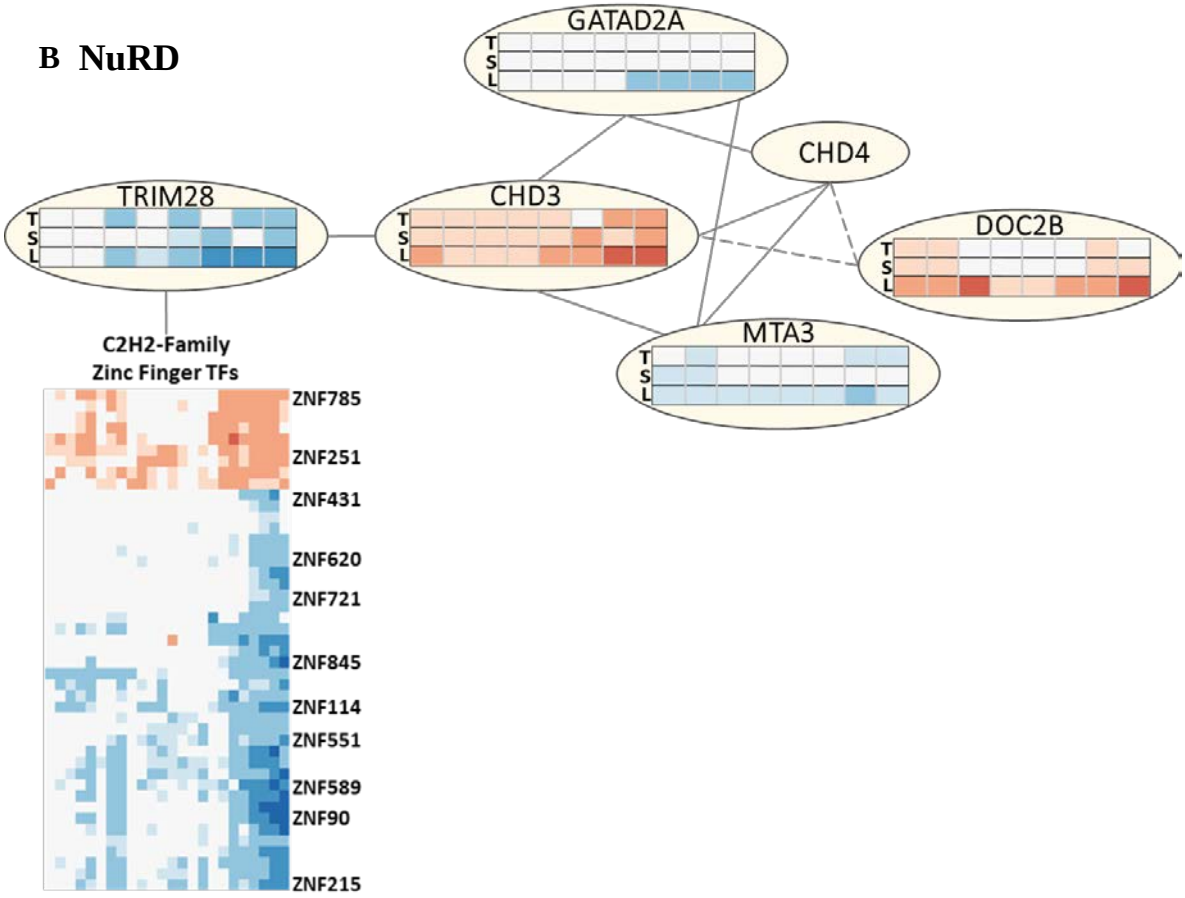
A. SWI/SNF class chromatin remodelers have core ATP-ase components SMARCA2 and SMARCA4. Interestingly, SMARCA4 is not differentially expressed, but SMARCA2 is DE in a cell-line dependent fashion, as are several other components.
B. NuRD chromatin remodeler has core ATP-ases CHD4 (not DE) and CHD3 (DE). TRIM28 is DE in a cell-line dependent fashion, and interacts with KRAB ZNF class transcription factors, many of which are likewise DE.

Lines between nodes are known protein-protein interactions from Stringdb; dashed lines are interactions postulated from the literature.

**A SWI/SNF**




**B NuRD**



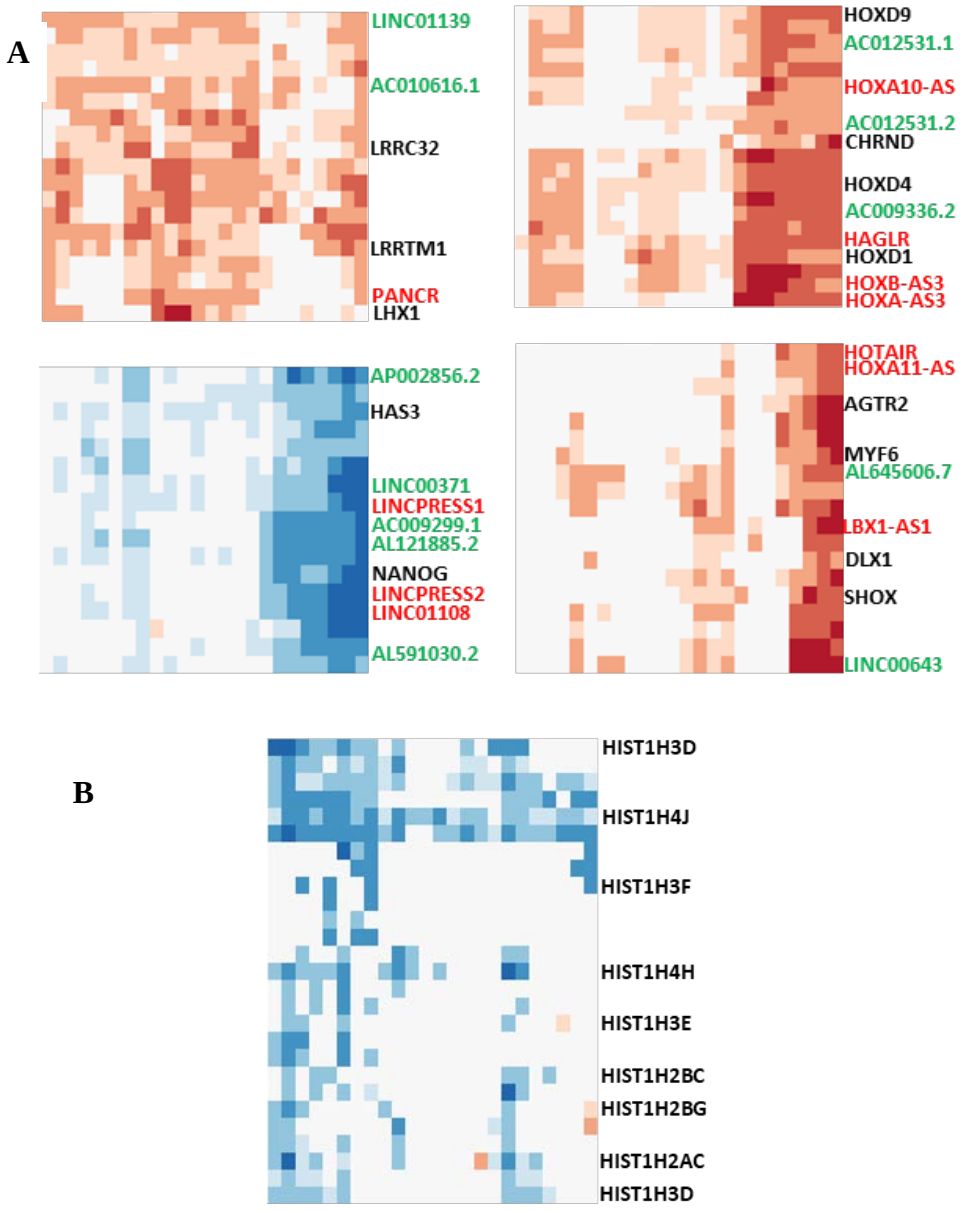


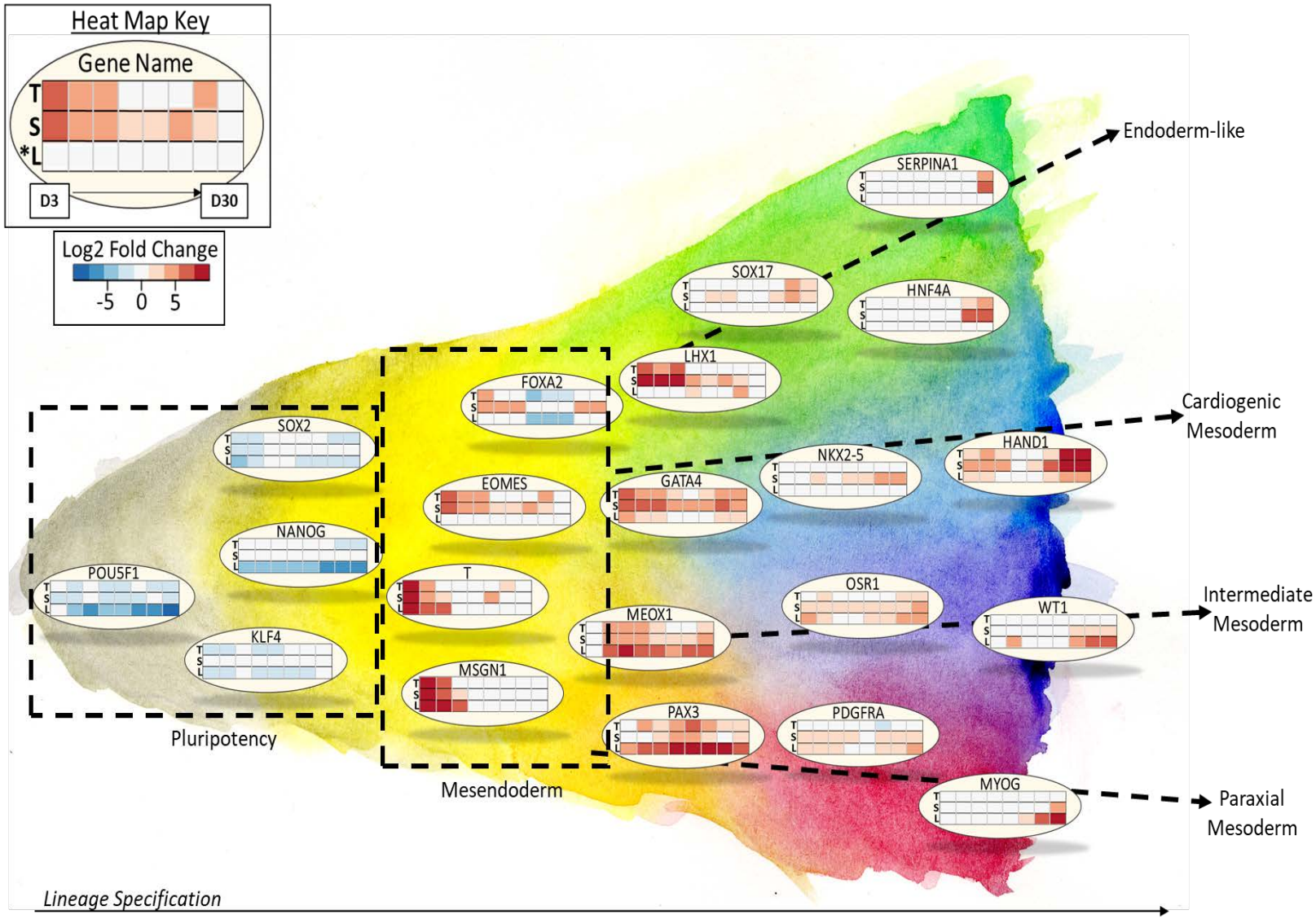

**Figure 3.6: Cell-line dependent expression of lncRNAs and HIST1H linker histones**

A. From top left, clockwise: (Red: known lncRNAs; Green: unannotated transcripts) 1.PANCR (PITX6-related lncRNA) clusters closely with LHX1 a key TF in endoderm specification. 2.HOX-associated lncRNAs cluster with the HOX genes, and several unannotated transcripts. 3.LINCPRESS1/2 and NANOG cluster with numerous unannotated transcripts. 4.HOTAIR, HOX11-AS cluster with Myf6, Shox and several unannotated transcripts.

B. Linker histone H1 components are DE in a cell line dependent pattern.

**Figure 3.7: Upstream expression of TFs that control cell line dependent endoderm or mesoderm commitment leads to downstream expression of derivatives from each respective germ layer**

The cell lines have variable expression of the core pluripotency regulators, as well as temporal differences in the expression of T (brachyury), the pan-primitive streak mesendoderm transcription factor. Notably, the L line has longer brachyury expression. By contrast the S and T lines have EOMES expression. S and T lines express FOXA2 and GATA4 early in differentiation, indicating endodermal fate. In addition to longer T expression, L also has longer MSGN1 expression, followed by stronger upregulation of Pax3 indicating more somitic mesodermal fate. L line ultimately has expression of MyoG (myogenesis), and WT1 (kidney, intermediate mesoderm derivative), while S and T lines have more HAND1 (cardiac) and HFN4A (liver) expression.


Heat Map Key
Gene Name
T
S
*L
D3
D30
Log2 Fold Change
-5 0 5
SERPINA1
Endoderm-like
SOX17
HNF4A
LHX1
FOXA2
SOX2
Cardiogenic
Mesoderm
NKX2-5
HAND1
EOMES
GATA4
NANOG
POU5F1
T
OSR1
Intermediate
Mesoderm
WT1
MEOX1
KLF4
MSGN1
PAX3
PDGFRA
MYOG
Paraxial
Mesoderm
Pluripotency
Mesendoderm
Lineage Specification

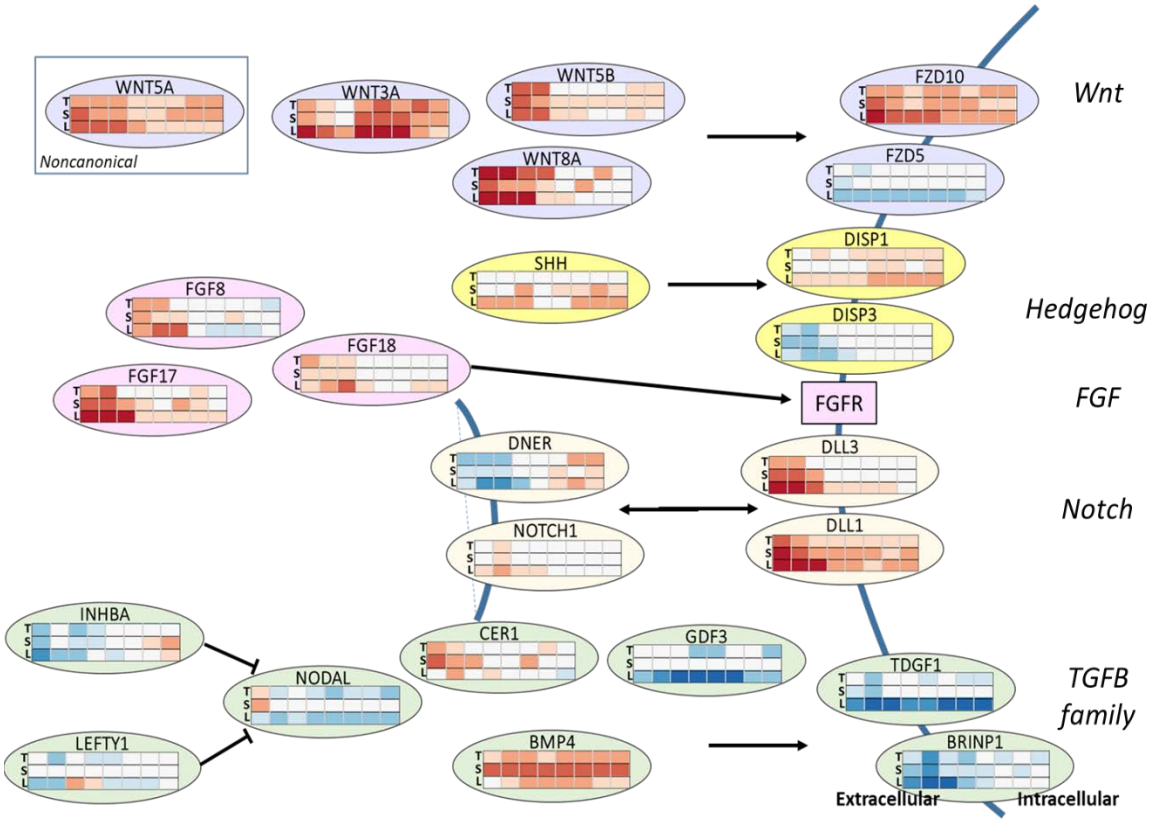



**Figure 3.8: Cell line dependent differential expression of several components of major signaling pathways**

Several Wnts are more strongly unregulated in the L line, including Wnt3A, 5B and 8A, as is the receptor FZD10. Likewise, Activin/Nodal and BMP4 are DE between cell lines. The coreceptor TDGF1 (Cripto) is strongly downregulated in the L line. Delta ligands DLL1 and DLL3 of the Notch pathway are upregulated longer in the L line, a sis NOTCH1 and its intracellular gene LFNG. Finally several FGFs (FGF8/17/18) exhibit the same pattern of expression between the cell lines: upregulated more strongly and for a longer duration in the L line compared to the S and T lines.

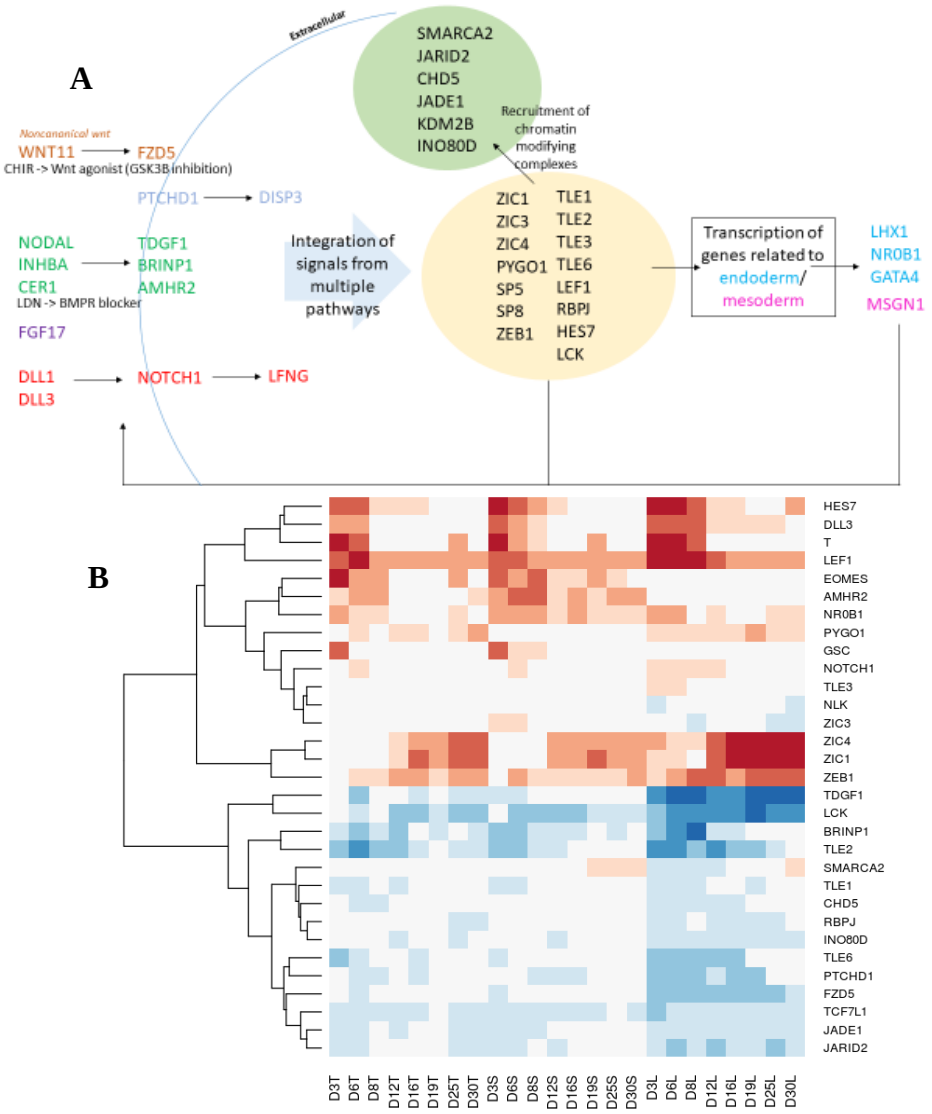



**Figure 3.9: Assessment of overall determinants of lineage specification, with attention to transcriptional cofactors/repressors of B-catenin**

A. Transcription factors downstream of signaling pathways integrate information by forming complexes with other transcription factors and chromatin modifying complexes to affect gene expression. In particular the transcriptional role of B-catenin is well-known in various contexts, including pluripotency and differentiation.

B. Expression profiles of candidate genes for knockdown. We have included several B-catenin transcriptional cofactors, as well as components of chromatin modifying complexes that have a particular expression pattern across all three cell lines. Several genes were selected whose knockdown is hypothesized to attenuate mesoderm commitment, as well as genes whose knockdown might enhance mesoderm commitment.

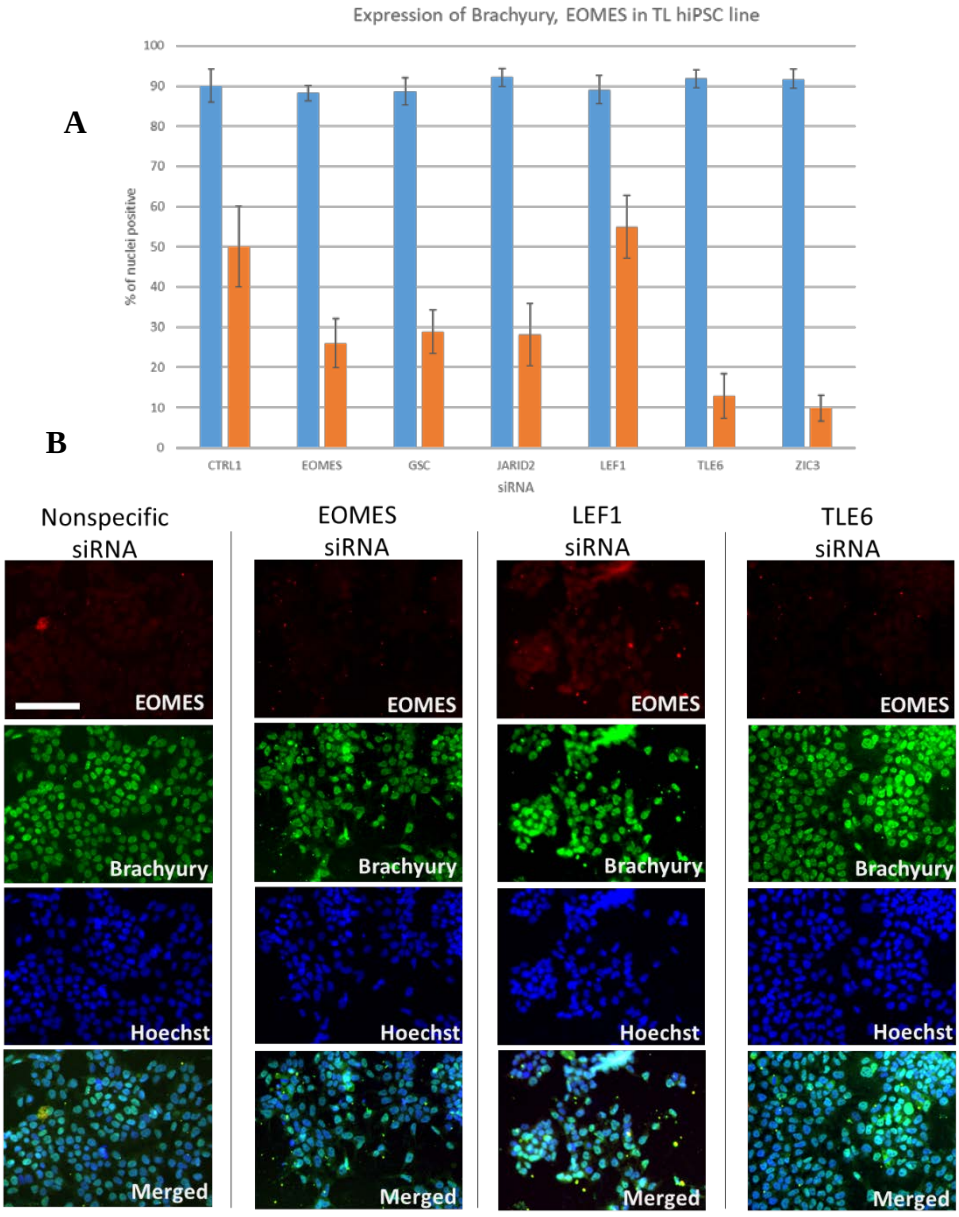



**Figure 3.10: Knockdown of TLE6, JARID2, ZIC3 leads to attenuated EOMES, but not brachyury, expression**

A. Knockdown of indicated genes by siRNA lead to reduced EOMES expression, but not brachyury expression. Eomes knockdown led to reduced EOMES expression (positive control) compared to scrambled control siRNA (negative control). GSC, another early mesendoderm marker had a similar effect on Eomes, as expected. Three candidate genes (JARID2, TLE6, ZIC3) for enhanced mesoderm expression had a similar effect, while LEF1 knockdown led to comparable or slightly higher EOMES expression than control.

B. IF staining for brachyury and EOMES

## Methods

### Pluripotent iPSC culture

iPSC lines were derived from peripheral blood cells. Pluropotent hiPSC colonies were maintained in feeder free conditions. Plates were coated with Matrigel (Corning) according to manufacturer's instructions and mTeSR1 (STEMCELL Tech) maintenance media changed daily was used to grow colonies. Cells were detached for replating as colonies using ReLeSR according to manufacturer's instructions. Cells were seeded as single cell by detaching with Versene (0.5 mM EDTA solution) and replated in mTeSR for ~24 hours in the presence of 2uM TZV before switching to differentiation medium

### Myogenic differentiation

Cells seeded as single cells on Matrigel coated plates at a seeding density of 20,000-30,000 cells per square centimeter. Five differentiation media conditions were used over the course of 35 days of myogenic differentiation. Days 1-3: DMEM/F12 was mixed with 1% NEAA, 1% Glutamax, 1% pen/strep, 1% ITS supplement, CHIR99021 (3uM), LDN-193189 (0.5uM). Days 4-6: DMEM/F12 was mixed with 1% NEAA, 1%Glutamax, 1% pen/strep, 1%ITS supplement, CHIR99021 (3uM), LDN-193189 (0.5uM), FGF (20ng/mL). Days 7-8: DMEM/F12, 15% KOSR (v/v), 1% NEAA, 1% Glutamax, 1% pen/strep, HGF (10 ng/mL), IGF (2ng/mL), FGF (20ng/mL), LDN (0.5uM), beta mercaptoethanol (0.1mM). Days 9-12: DMEM/F12, 15% KOSR, 1% NEAA, 1%Glutamax, 1% pen/strep, IGF (2ng/mL), beta mercaptoethanol (0.1mM). Days 13-35: DMEM/F12, 15% KOSR, 1% NEAA, 1%Glutamax, 1% pen/strep, HGF (10 ng/mL), IGF (2ng/mL), beta

mercaptoethanol (0.1mM). Media changed daily for the first 12 days, followed by half media change every day for the remainder of the differentiation

Immunofluorescent staining

Immunofluorescent staining was performed using the following primary antibodies: PAX3, PAX7, MF20 (1:200; Developmental Studies Hybridoma Bank), MYF5 (1:200; Santa Cruz Biotechnology), desmin (1:200; Abcam), and human lamin A/C (1:50; Vector Laboratories). The following secondary antibodies were used: goat anti-rat Alexa 546 (1:200; Life Technologies), goat anti-mouse Alexa488 (1:250; Life Technologies), and goat anti-rabbit Alexa 546 (1:200; Life Technologies). For immunofluorescent staining of cells grown on tissue culture plates, cells were fixed in 4% PFA for 10 min at room temperature. Immediately before staining, the cells were permeabilized with 0.1% (v/v) Triton X-100 and blocked with 3% (w/v) BSA for 30 mins. Cells were stained with primary antibodies diluted in 1% BSA overnight at 4°C, washed 3 times with PBS, and incubated with secondary antibodies for 1 hr at room temperature. The nuclei were stained with Hoechst 33342 (2 µg/ml; Life Technologies) for 5 mins at room temperature. For immunofluorescent staining of tibialis anterior (TA) muscles, samples were first embedded in optimal temperature cutting compound (OCT) for cryosectioning and sections (having around 20 µm thickness) were fixed with 4% PFA for 10 mins at room temperature. Next, sections were permeabilized with 0.3% Triton X-100, blocked with 3% BSA for 1 hr at room temperature, and stained with human lamin A/C. Imaging was performed using a fluorescence microscope (Carl Zeiss; Axio Observer A1).

RT-qPCR

hiPSCs cultured for various amount of time according to the myogenic differentiation protocol were examined for changes in the gene expression as a function of culture time. RNA was extracted from cell cultures using TRIzol according to the manufacturer's instructions. For each sample, 1 μg of RNA was reverse-transcribed to complementary DNA (cDNA) using an iScript cDNA synthesis kit (Bio-Rad, catalog no. 170-8891). Real-time PCR reactions were run on ABI Prism 7700 Real-time PCR Cycler (Applied Biosystems).For qPCR analysis of selective genes, SYBR Select Master Mix (Life Technologies, catalog no. 4472908) was mixed with various primers (Pax7, Myf5, MyoD, MyoG). The expression of each target gene was normalized to that of corresponding 18S, a housekeeping gene. The expression levels were normalized to that of undifferentiated, pluripotent hPSCs and presented as log2(fold change).

RNA extraction, quality assessment, library preparation, and sequencing

RNA samples were collected for ten time points with two biological replicates for each of three cell lines using TRIzol according to manufacturer's instructions. At the UCSD Institute for Genomic Medicine, RNA RIN scores for each sample were calculated using Agilent 4200 TapeStation instrument. Library prep carried out with the Illumina TruSeq stranded mRNA kit. Samples were sequenced on an Illumina hiSeq4000. 100 bp paired end reads with about 30 million reads per sample were generated.

RNAseq data quality check

The quality of raw reads was assessed using FastQC (Babraham Bioinformatics). FastQC reads for multiple samples were compiled using MultiQC (Phil Ewels).

Read alignment and differential expression

Alignment to the human transcriptome (Ensembl version GRCh38.p10) that consisted of cDNA and ncRNA transcripts was preformed using Salmon. Summation of transcript read counts to gene counts done using tximport. Differential gene expression analysis was carried out using DeSeq2. Preliminary downstream analyses were carried out using R software for statistical computing.

## References


Abady, S., Malcuit, C., Kashpur, O., Kole, D., Holmes, W.F., Hedblom, E., Page, R.L., Dominko, T., 2010. Expression of NANOG and NANOGP8 in a variety of undifferentiated and differentiated human cells. Int J Dev Biol 54, 1743–1754.

Borchin, B., Chen, J., Barberi, T., 2013. Derivation and FACS-Mediated Purification of PAX3+/PAX7+ Skeletal Muscle Precursors from Human Pluripotent Stem Cells. Stem Cell Reports 1, 620–631.

Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells, 2012a. . Cell Stem Cell 10, 440–454.

Goodman, F.R., 2003. Congenital abnormalities of body patterning: embryology revisited. Lancet 362, 651–662.

Kempf, H., Olmer, R., Haase, A., Franke, A., Bolesani, E., Schwanke, K., Robles-Diaz, D., Coffee, M., Göhring, G., Dräger, G., Pötz, O., Joos, T., Martinez-Hackert, E., Haverich, A., Buettner, F.F.R., Martin, U., Zweigerdt, R., 2016. Bulk cell density and Wnt/TGFbeta signalling regulate mesendodermal patterning of human pluripotent stem cells. Nature Communications 7, 13602.

Kotoula, V., Papamichos, S.I., Lambropoulos, A.F., 2008. Revisiting OCT4 expression in peripheral blood mononuclear cells. Stem Cells 26, 290–291.
Li, M., Belmonte, J.C.I., 2017. Ground rules of the pluripotency gene regulatory network. Nature Reviews Genetics 18, 180–191.

Mendjan, S., Mascetti, V.L., Ortmann, D., Ortiz, M., Karjosukarso, D.W., Ng, Y., Moreau, T., Pedersen, R.A., 2014. NANOG and CDX2 Pattern Distinct Subtypes of Human Mesoderm during Exit from Pluripotency. Cell Stem Cell 15, 310–325.

Mills, K.M., Szczerkowski, J.L.A., Habib, S.J., 2017. Wnt ligand presentation and reception: from the stem cell niche to tissue engineering. Open Biol 7.

Nie, Z., Yan, Z., Chen, E.H., Sechi, S., Ling, C., Zhou, S., Xue, Y., Yang, D., Murray, D., Kanakubo, E., Cleary, M.L., Wang, W., 2003. Novel SWI/SNF Chromatin-Remodeling Complexes Contain a Mixed-Lineage Leukemia Chromosomal Translocation Partner. Mol Cell Biol 23, 2942–2952.

Rios, A.C., Serralbo, O., Salgado, D., Marcelle, C., 2011. Neural crest regulates myogenesis through the transient activation of NOTCH. Nature 473, 532–535.

Wang, Z., Oron, E., Nelson, B., Razis, S., Ivanova, N., 2012. Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. Cell Stem Cell 10, 440–454.

Zangrossi, S., Marabese, M., Broggini, M., Giordano, R., D'Erasmo, M., Montelatici, E., Intini, D., Neri, A., Pesce, M., Rebulla, P., Lazzari, L., 2007. Oct-4 expression in adult human differentiated cells challenges its role as a pure stem cell marker. Stem Cells 25, 1675–

# Chapter 4: Functional validation of hiPSC-derived myogenic progenitors on an in vitro skeletal muscle on a chip platform and an in vivo murine model of muscular dystrophy

## Introduction

Harnessing the ability of stem cells and their derivatives to promote regeneration of compromised tissues could be of great potential to reinstating the functionality of dysfunctional tissues/organs. Cell based therapies hold great promise to treat different skeletal muscle defects ranging from traumatic injuries to age-related and genetic muscle wasting diseases such as muscular dystrophies. The transplanted cells could contribute to tissue repair by either differentiating into tissue specific cells or by the secretion of trophic factors. Incorporation, differentiation, and integration of the healthy donor cells within the multinucleated myotubes could compensate for the "diseased" cells and could contribute to improving the tissue function. On the other hand, secretion of trophic factors could ameliorate the disease pathology by rejuvenating the host tissue environment. Studies over the years have shown that while cell transplantation could contribute to skeletal muscle tissue regeneration/repair, this approach suffers severely from low viability and poor-to-modest engraftment of transplanted cells. In addition to cell-based therapy, stem cell-derivatives (especially hiPSC-derived cells) could be used to create powerful technological platforms such as patient- and disease- specific *in vitro* tissue surrogates. In this chapter, I will validate the potential of hiPSC-derived cells to form *in vitro* skeletal muscle tissues

and contribute to *in vivo* tissue repair upon transplantation, as well as address the some of the associated bottlenecks.

One of the limitations of hiPSCs as a cell source for regenerative medicine (which involves cell based therapy and cell-based technological platforms) is their heterogeneous differentiation. Hence, developing or identifying experimental conditions that yield efficient derivation of myogenic progenitor cells from pluripotent hiPSCs is a prerequisite. There exist a few experimental protocols describing derivation of myogenic progenitor cells with varying outcome. We have adapted an existing protocol (selected after an extensive comparison) and modified it to derive robust myogenic progenitor cells from multiple hiPSC lines: three healthy lines, and two lines derived from patients with juvenile dermatomyositis. We will examine experimental condition-dependent cell commitment to the myogenic lineage for multiple myogenic markers as a function of time. The experimental conditions will be used to derive myogenic progenitors with varying extents of differentiation commitment- either early- (MYF5 positive) or late- (MYOG and PAX7 positive) commitment for downstream applications.

Stage specificity or extent of differentiation commitment of the hiPSC-derived myogenic progenitors could play an important role in determining their ability to form functional 3D skeletal muscle tissues. Using the early- and late- extent of differentiation commitment myogenic progenitor cell populations, we will test the cell phenotype-dependent in vitro tissue formation ability of hiPSC-derived cells by using a skeletal muscle-on a-chip platform developed in our lab. Additionally, we will test the functionality

of these microtissues with respect to their response to physiological mechanical and chemical stimuli.

Finally, we will validate the hiPSC-derived myogenic progenitors in an in vivo context. Potential therapeutic application of hiPSC-derived cells relies on their ability to survive, differentiate into tissue specific cells, and integrate with the existing host tissue *in vivo* without contributing to teratoma formation. Both the phenotype of the donor cells as well as the host tissue environment play a pivotal role in determining the *in vivo* function of the transplanted cells. Here we test the early- and late- specified progenitors for their survival and function in a murine model of muscular dystrophy.

## Results

### Comparison of two published protocols for the induction of myogenic differentiation of hiPSCs

We compared two protocols (Borchin, 2013; Chal, 2016) (Figure 4.1) that utilize small molecules and growth factors to induce the myogenic differentiation of hiPSCs to determine which gave the most consistent myogenic differentiation for multiple cell lines in the shortest amount of time. Both protocols rely on CHIR99021 for initial paraxial mesoderm induction, followed by FGF2-mediated expansion of the myogenic progenitor population. The final, longest stage of differentiation in both protocols withdraws FGF2 to induce maturation of the myogenic population. In addition, one of the protocols (Chal et al) also utilizes BMPR inhibition via LDN-193189 early in differentiation. This protocol

also adds HGF and IGF signaling, as well as a KOSR-containing medium in the latter stages of differentiation.

In both protocols, cells reach confluence within the first 4-6 days of differentiation and can be seen to form dense compactions that protrude from the monolayer within the first 10 days of differentiation. In Chal et al, (Figure 4.2) these dense areas are more numerous and in general arise earlier across multiple cell lines than under culture conditions specified in Borchin et al. (Figure 4.3) Nuclear staining with Hoescht shows these mounds to be comprised of cells that have grown on top of each other. With prolonged differentiation, these areas continue to enlarge. There are some variations in cell-line to cell-line morphology in terms of the size of these cell populations.

We characterized the temporal progression and extent of myogenesis under both culture conditions using the cell line that exhibited the most robust myogenic differentiation over numerous attempts at differentiation using both protocols (Figure 4.4). When the hiPSCs were cultured according to the protocol by Borchin et al, we observed early expression of Myf5 by mid differentiation and expression of desmin by day 25 onwards. However, we did not detect any expression of Mf20, MyoG or Pax7 in 30 days of differentiation. In the protocol by Chal et al, we also observed early expression of Myf5 and desmin expression by day 19. However, we were also able to detect by immunofluorescence Mf20, MyoG and Pax7 at day 25, with increased expression of these myogenic markers at day 30. Our results indicate that the protocol by Chal et al leads to more robust myogenic differentiation within 30 days compared to the protocol by Borchin et al.

Downstream modification for selective expansion and terminal differentiation of myogenic progenitor cells

The dense, raised regions preferentially give rise to cell populations that stain positive for markers of myogenesis. Using the protocol by Chal et al, these dense regions contain cell populations that stain positive for myosin heavy chain (Mf20) and MyoG, as well as Pax7, a marker for satellite cells, the endogenous stem cells of skeletal muscle. However, these areas of robust myogenesis are surrounded by cell populations of that do not stain positive for the aforementioned markers of myogenesis (Figure 4.5). This indicates that 30 days of differentiation according to the published protocols give rise to a heterogeneous population. We have modified and optimized the existing protocols for the selective expansion and terminal differentiation of myogenic progenitor cells and successfully tested this protocol using cells that were differentiated for an initial 30 days according to both protocols by Chal et al and Borchin et al.

Two growth media were compared for passaging and expanding the myogenic progenitor cells: “standard” growth medium consisting of DMEM+10% FBS and a “modified” growth medium consisting of DMEM/F12+20%KOSR+FGF. The addition of FGF promotes the proliferation of myogenic cells. Additionally, two differentiation media were tested for terminal differentiation: the first contained 2% horse serum, which is a myogenic differentiation medium that has been widely used in the literature for this purpose, while the second contained ITS supplement, which has been used more recently for myogenic differentiation. After 30-35 days of differentiation, cells that were cultured according to both protocols were passaged into either standard or modified growth

medium. They were allowed to proliferate and reach 70-80% confluency before switching to either 2%HS or ITS differentiation medium. Each growth medium was tested in combination with both of the differentiation media, for a total of four combinations. (Figure 4.6) Cells proliferated in both types of growth media and exhibited varying extents of myogenic differentiation in both differentiation media. The most robust, homogeneous myogenic differentiation, as gauged by immunofluorescent staining for MF20 and desmin, was achieved by the combination of expansion in modified growth medium followed by terminal differentiation in ITS-supplemented differentiation medium. This combination of media resulted in relatively homogeneous myogenic differentiation for all cell lines tested. The other three combinations of growth and differentiation media led to heterogeneous differentiation, with only few long cells positive for Mf20 and desmin, surrounded by numerous, small round cells that were negative for myogenic markers.

Figure 4.6 shows the outcome of the various media combinations using cells that were initially differentiated for 30 days pre-passage according to the Borchin et al protocol. Figure 4.7 shows a similar result using cells that were differentiated for an initial 30 days using the Chal et al protocol. Although the protocol by Borchin et al did not result in cells that were positive for Mf20 in 30 days of initial differentiation, we were able to achieve Mf20 positive cells using our protocol to passage, expand and terminally differentiate the cells beyond 30 days. Similarly, we achieved a relatively homogeneous population of long, multinucleated Mf20 positive cells when we applied our protocol beyond 30 initial days of differentiation according to the protocol by Chal et al. (Figure 4.8). Since we were able to

achieve better and more consistent overall differentiation results using the protocol by Chal et al, we used this protocol in our subsequent experiments.

Freeze/thaw and expansion of population by serial passaging

The ability to freeze and thaw cells, as well as the ability to expand of the population of myogenic progenitors would greatly increase the feasibility of downstream applications such as creating *in vitro* tissue models and cell transplantation that may require large numbers of cells. After passaging cells at 30-35 days of initial differentiation and expanding them for several days in growth medium, the cells can be frozen and later thawed successfully into the same modified growth medium. The cells can be differentiated as normal in differentiation medium.

To test whether the thawed cells retain their myogenic nature with serial passaging, we serially passaged cells up to four passages. At each passage, when the cells reached 70-80% confluence in growth medium, they were either switched to differentiation medium, or passaged again for further expansion in growth medium. (Figure 4.9) qPCR of Pax7, MyoD, and MyoG shows upregulation of all three markers in P1 + 7 days of differentiation when compared to P1 + 0 days of differentiation (at the end of expansion in growth medium, indicating an increase in myogenic maturation as expected. We noted retention of myogenic potential up to four passages, as MyoD and MyoG expression increase after 7 days in differentiation medium even after four passages. This indicates the presence of maturing myogenic progenitors, although differentiation is blunted compared to early passage. Notably however, Pax7 expression is attenuated by P4. Altogether, these data

indicate that while cells can be passaged and expanded, it is preferable to use early passage cells.

Derivation of cell populations with early- and late- extent of differentiation commitment

Since extent of myogenic differentiation commitment could affect cell function in downstream applications such as creation of *in vitro* tissue models and cell transplantation, we derived two distinct cell populations that differ in their expression of early- or late-myogenic regulatory transcription factors. (Figure 4.10 A-D) After passaging the cells at 30-35 days of initial differentiation, they are cultured in differentiation medium for less or more time to induce early- or late- commitment, respectively. The early specified group is predominantly uninucleated and expresses Myf5; a handful of cells express desmin as well. The late specified group comprises of a mixed population of myogenic progenitors, including long, multinucleated Myf5, MyoG, Mf20, dystrophin positive cells. Alongside these cells are uninucleated Pax7 positive cells. To see whether MyoG precedes Pax7 expression or vice versa, we characterized cells in between the early- and late- specified groups. In this middle-specified population, we observed an increased number of long, thin uni- or- bi-nucleated cells positive for MyoG and Mf20; however, we did not detect Pax7 positive cells. This indicates that MyoG positive cells arise before satellite-like cells in culture, and may play a role in establishing an in vitro niche for Pax7 positive cells.

In vitro validation of functional hiPSC-derived myogenic progenitors using a skeletal muscle-on-a-chip platform

Ongoing work in our lab has led to the development of a skeletal muscle-on-a-chip platform for the in vitro formation of muscle microtissues. The chip platform offers several

benefits for tissue formation in addition to 3D culture: The pillars that anchor both ends of the tissue allow for compaction and alignment, which might lead to more efficient myogenic differentiation. (Figure 4.11A) Furthermore, the ends of the pillars are attached to a deformable acrylamide gel with embedded fluorescent microbeads. Thus, contractile forces and passive tension exerted by the microtissue on the pillars can be measured. This offers a functional readout for tissue formation. Figure 4.11B shows maturation of the microtissue from 3 to 7 days, as evidenced by the increase in the number of Mf20/Desmin positive cells. At 7 days of differentiation, microtissues express both MyoG and Pax7 positive cells.

We tested both early- and late- specified populations to see whether one could more effectively form a 3D tissue, but we did not note any significant difference. As both populations were readily able to form tissues in vitro, we used the late population for the practical consideration of increased cell number, as the cells proliferate somewhat during differentiation. We next applied a periodic loading to the microtissues to gauge their response to passive mechanical stimulation. We applied the loading in a dose-dependent manner by varying the number of times the microtissues were stimulated: the control group received no stimulation, another group received compressions at long-intervals once every 6 hours, and a third group received stimulation at short intervals three times every six hours. After three days of stimulation, the group that received the most mechanical loading had the most expression of MyoG. Next we tested the response of the microtissues to the neurotransmitter acetylcholine. In vivo, acetylcholine released from motor neurons stimulates contraction in skeletal muscle cells via acetylcholine receptors located on the

skeletal muscle cell membrane at the neuromuscular junction. As our system does not have motor neurons, we flowed media supplemented with exogenous acetylcholine into the chip to test whether it would stimulate the myogenic microtissues to contract. We observed contraction in the microtissues over about a five to ten minute period, and subsequent washing out of the acetylcholine caused some tissues to return to their original state. However, we also noted that some tissues snapped, as they were unable to withstand the forces generated by the contracting cells. Taken together, the two experiments show that the hiPSC-derived myogenic microtissues have functional responses to physiological stimuli: they have hastened maturation in response to mechanical loading and contract in response to the neurotransmitter acetylcholine.

In vivo validation of hiPSC-derived myogenic progenitors in a murine model of muscular dystrophy

Finally, we tested the ability of hiPSC-derived myogenic progenitors to survive and function when transplanted into a murine model of muscular dystrophy. The transplanted cells survive up to 28 days post-transplantation, and give rise to dystrophin-positive fibers. (Figure 4.13) Human lamin A/C positive cells were observed within the host myotubes, indicating that the transplanted hiPSC-derived cells had fused with host cells, and were able to function within the host tissue.

## Discussion

In this chapter, we established a protocol, modified from the literature, for the generation of a purified population of hiPSC-derived myogenic progenitors. We first

compared two published protocols for the consistency with which they induced myogenic progenitors that expressed markers of terminal specification from multiple cell lines. The protocol by Chal et al more consistently produced cells that expressed Mf20, MyoG, and Pax7 by 30 days of differentiation, while the protocol by Borchin et al had cells that were positive only for desmin. However, neither protocol had homogeneous expression of these myogenic markers. In order to purify, expand, and terminally specify the myogenic progenitor population, we modified the published protocols through passaging the cells into a 20% KOSR growth medium, supplemented with FGF, followed by an ITS-supplemented medium for terminal differentiation. Cells derived with both the Borchin and Chal protocols were cultured in this way to produce a relatively purified population of Mf20-positive cells, with numerous multinucleated myotubes. That we were able to achieve this result starting with either protocol demonstrates that a relatively inefficient differentiation can be at least partially recovered by purification and selective expansion of myogenic progenitors; this is a prerequisite for downstream in vitro and in vivo applications.

The derivation of populations specified to the myogenic lineage but distinct in their expression of MRFs related to terminal differentiation-namely MyoG and Pax7- could be of utility in downstream applications. In particular, the early-specified population contains predominantly uninucleated Myf5 positive cells. Several of these cells are desmin positive, but are largely Mf20, dystrophin and MyoG negative. Furthermore, there are few to no Pax7 positive cells detectable by immunofluorescent staining. With increasing time of differentiation, this early-specified population gives rise to long, multinucleated cells that

are Myf5, Mf20, dystrophin and MyoG positive. Furthermore, there are numerous uninucleated Pax7 positive cells located adjacent to the long multinucleated cells. It is interesting to note that the satellite-like Pax7 positive cells appear after the appearance of MyoG positive cells. This indicates that terminally specified, MyoG positive cells may spontaneously create a niche in vitro for Pax7 positive cells, perhaps by the secretion of extracellular matrix proteins, or through cell-cell contact. Furthermore, it appears that the satellite-like cells in vitro can give rise to both MyoG+/Pax7- multinucleated cells, as well as uninucleated Pax7 positive cells up to 3 or 4 passages. This may be a result of asymmetric division of the Pax7 positive cells in vitro.

Since the use of early- or late-committed myogenic progenitors could affect outcomes in both in vitro and in vivo applications, we tested the ability of both cell populations to form 3D tissue in vitro. However, we did not observe any appreciable difference in the ability of the early vs late populations to form in vitro tissues in our skeletal muscle on a chip model. The skeletal muscle-on-a-chip platform allowed us to create functional hiPSC-derived myogenic progenitor microtissues. We validated two aspects of physiologically normal function using this platform: First, we took advantage of the microfluidic flow aspect of the system to apply passive mechanical loading to the tissues. We varied the frequency of compression from none to once every six hours to thrice every six hours to test whether less or more frequent stimulation had an effect on tissue maturation. We found that increased frequency of stimulation led to a hastening of tissue maturation, as evidenced by the dose-dependent increase in the number of MyoG positive nuclei. This strategy could be used as a tool to improve and accelerate differentiation in

vitro, which addresses an important constraint in the clinical utility of in vitro hiPSC-based platform. Finally, we observed contraction of the microtissues in response to exogenous acetylcholine. This physiological response to acetylcholine is an important benchmark to validate our differentiation strategy: To respond to acetylcholine, the cells should have functional acetylcholine receptors, should depolarize normally, and finally should have sufficiently developed sarcomeres to affect contraction. Taken together, the two tests of microtissue function in response to physiological stimuli validate our hiPSC differentiation strategy and highlight the potential use of the skeletal muscle on a chip platform for patient-specific modeling and for integration with other organ-on-a-chip platforms.

Similarly, the early- and late- specified cell populations may have different behaviors when transplanted in vivo, including survival and ability to migrate. Cells of both early and late populations survived for at least 14 days in a cardiotoxin-injured murine model of muscular dystrophy and gave rise to dystrophin positive fibers.

In summary, we have established a modified protocol for the purification, selective expansion, and terminal differentiation of hiPSC-derived myogenic progenitors. We have furthermore established an in vitro skeletal-muscle on a chip model that we used to validate physiologically relevant aspects of myogenesis- namely, hastening of maturation in response to passive compression, and contraction in response to stimulation by acetylcholine. Finally, we tested the ability of these cells to give rise to dystrophin positive cells in a murine model of muscular dystrophy.

**Acknowledgements**

Chapter 4 is a modified presentation of material that is being prepared for presentation as “Using human induced pluripotent stem cells to create a functional skeletal muscle-on-a-chip” by Agrawal, A, Nayak, P, Varghese, S. The dissertation author was co-author of this material.

**Figures**

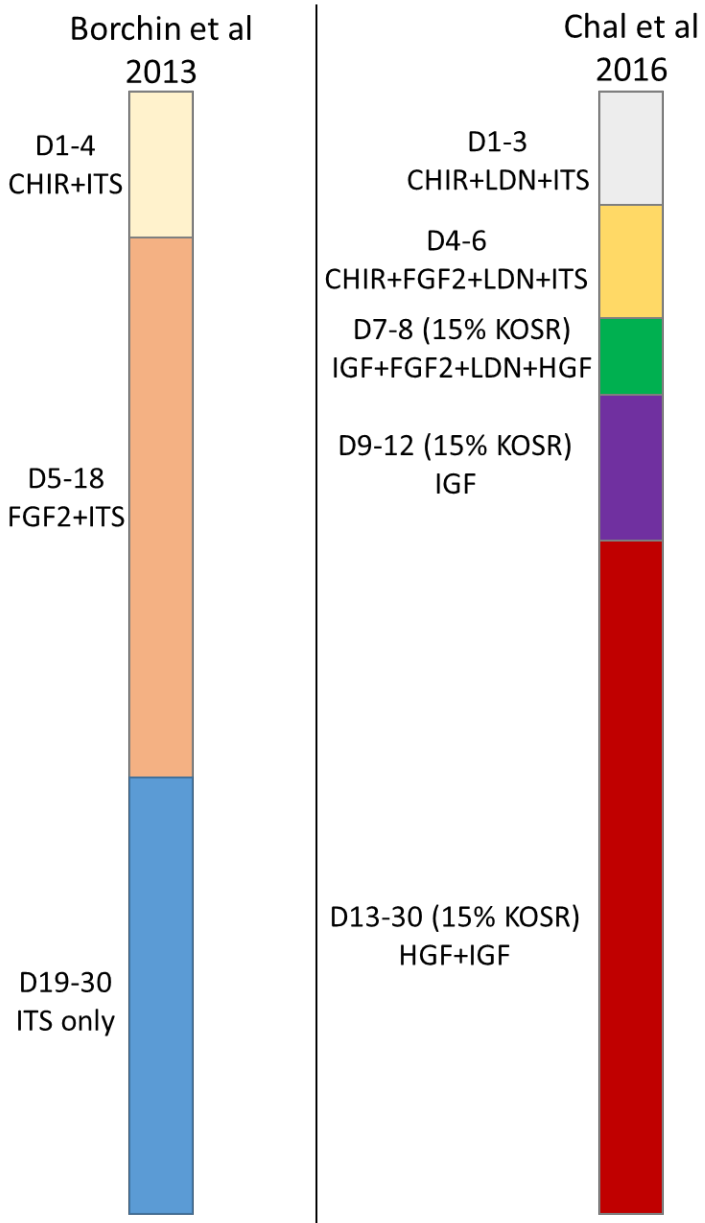



**Figure 4.1: Comparison of media composition of two myogenic differentiation protocols**

Two protocols for the myogenic differentiation of hPSCs using small molecule and growth factor cues only. CHIR=CHIR99021, GSK3B inhibitor; ITS= insulin, selenium, transferrin; LDN=LDN193189, BMP Type 1 receptor inhibitor; FGF2=fibroblast growth factor 2; IGF=insulin-like growth factor 1; HGF=hepatocyte growth factor; KOSR=knock-out serum replacement

**Figure 4.2: Brightfield images depict morphologies of 5 hiPSC lines over 35 days of myogenic differentiation according to culture conditions by Chal et al.**

Cells become confluent within 4 days and form dense regions that appear dark on brightfield; these regions expand with increasing time of differentiation. There is cell line to cell line variability in morphology, but the trend is consistent between cell lines. Each column represents a different cell line.

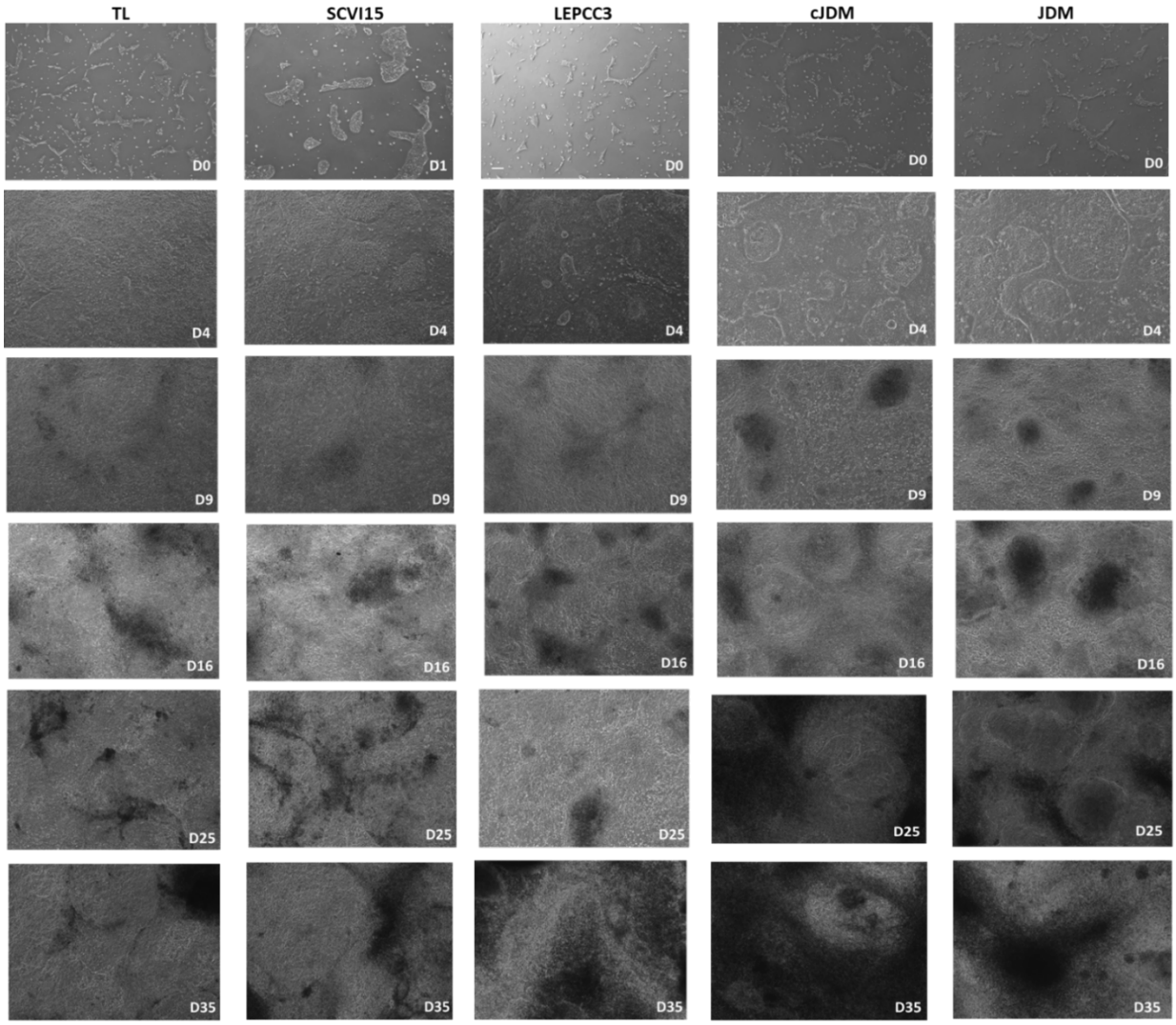
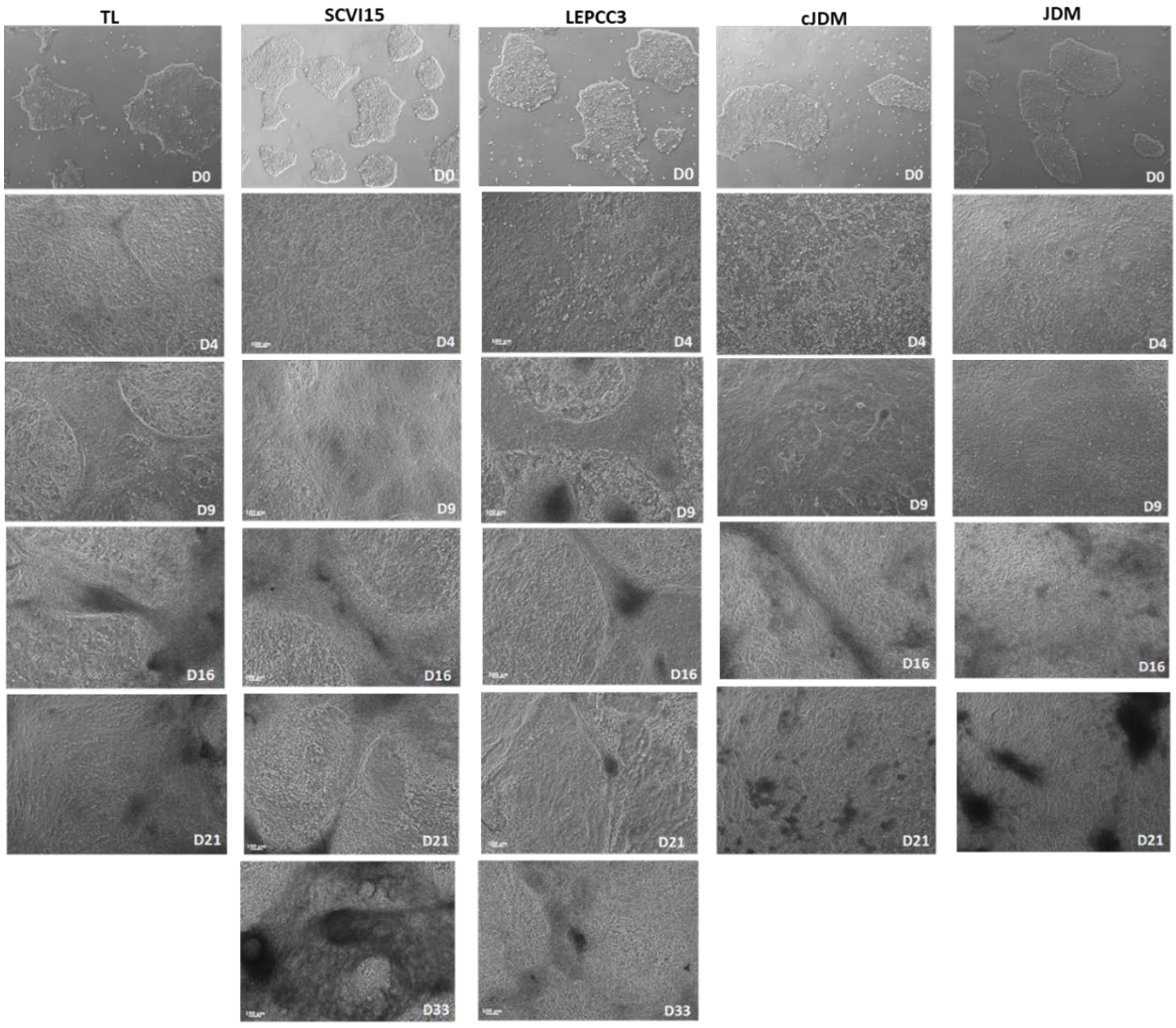

TL
SCVI15
LEPCC3
cJDM
JDM
D0
D1
D0
D0
D0
D4
D4
D4
D4
D4
D9
D9
D9
D9
D9
D16
D16
D16
D16
D16
D25
D25
D25
D25
D25
D35
D35
D35
D35
D35

**Figure 4.3: Brightfield images depict morphologies of 5 hiPSC lines over 30+ days of myogenic differentiation according to culture conditions by Borchin et al.**

Cells start as small pluripotent colonies and become confluent in differentiation medium within four days, similar to cells in culture conditions by Chal et al. As before, the cells grow into dark-appearing, dense regions. There is cell line to cell line variability in morphology, but the trend is consistent between cell lines. Each column represents a different cell line.


TL
SCVI15
LEPCC3
cJDM
JDM
D0
D4
D9
D16
D21
D33

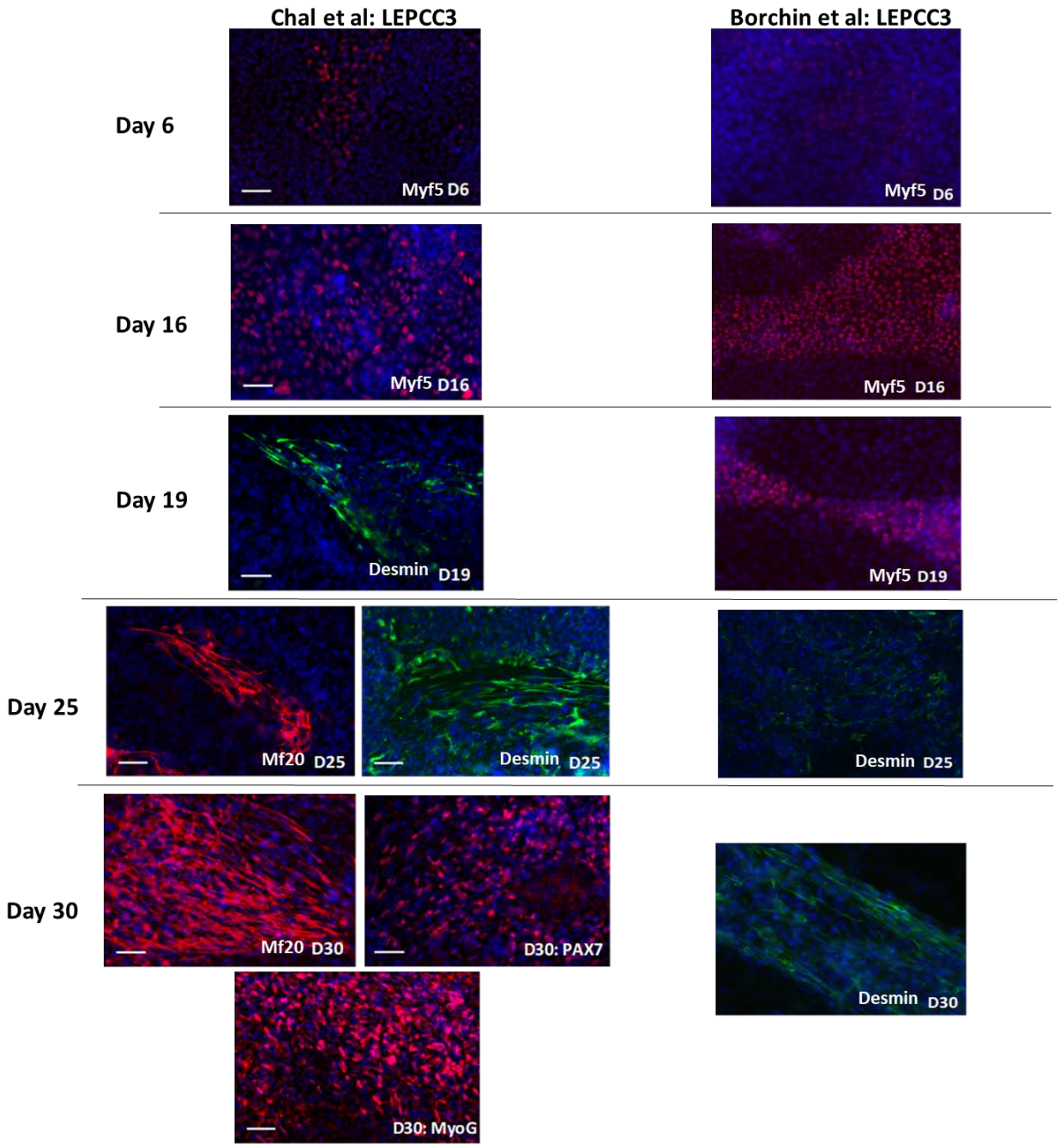
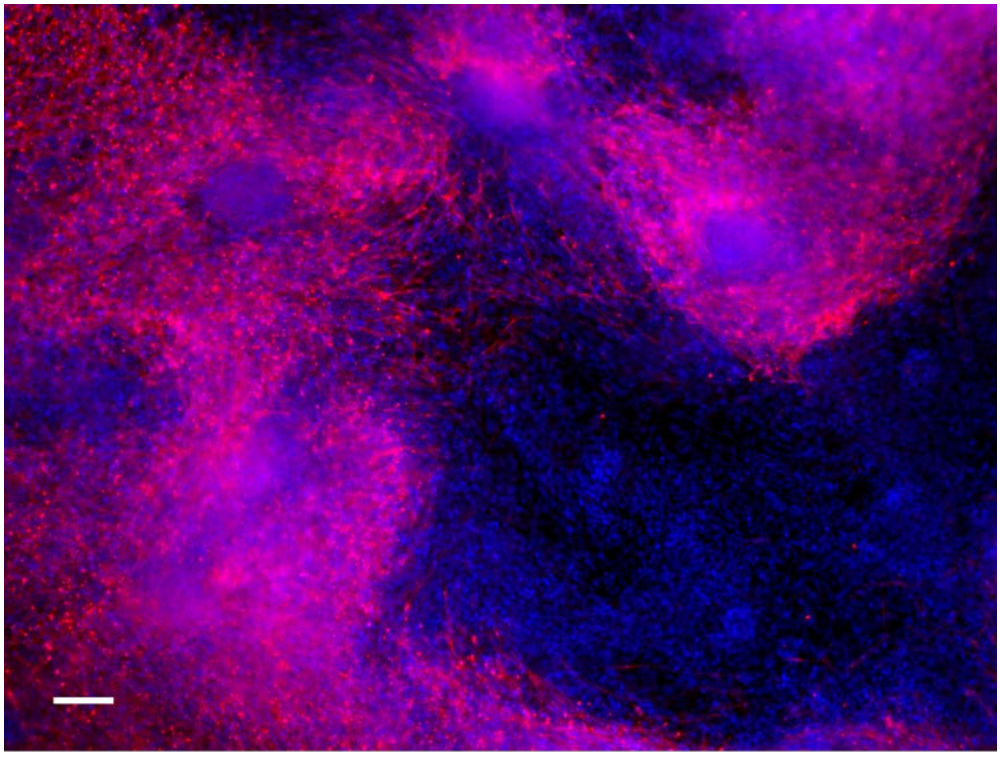



**Figure 4.4: Comparison of timing and extent of myogenic differentiation of hiPSCs cultures according to Chal et al and Borchin et al**

We stained for multiple markers of early and late myogenic specification to characterize the progression of myogenic differentiation of one cell line, using both protocols. Both protocols expressed Myf5 early, but the cells cultures with Chal et al expressed desmin by day 19. By contrast, cells grown with Borchin et al only expressed desmin at day 25. Chal et al conditions gave rise to Mf20, MyoG, and Pax7 positive cells by 30 days, while Borchin et al did not. Scale bar=100um

A



Mf20/Hoechst

B

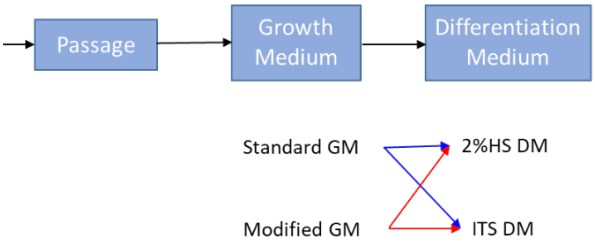



**Figure 4.5: Heterogeneous population at day 30 of differentiation necessitates a strategy for the purification of myogenic progenitors**

A. 5x image of immunofluorescent staining for Mf20 at day 30 of differentiation using the Chal et al protocol shows myogenic progenitor populations that stain positive for Mf20, and populations that do not, indicating heterogeneity during differentiation. Scale bar =100um

B. Schematic shows procedure for optimization of culture conditions for the purification, selective expansion and terminal differentiation of the myogenic population. We first passage the cells at low density into a growth medium containing FGF2, or a standard growth medium without FGF2. Then we switch both growth medium conditions to either a serum-free differentiation medium containing ITS supplement, or a 2% horse serum differentiation medium, for four total experimental conditions. GM = growth medium; HS = horse serum; ITS = insulin, selenium, transferrin; DM= differentiation medium

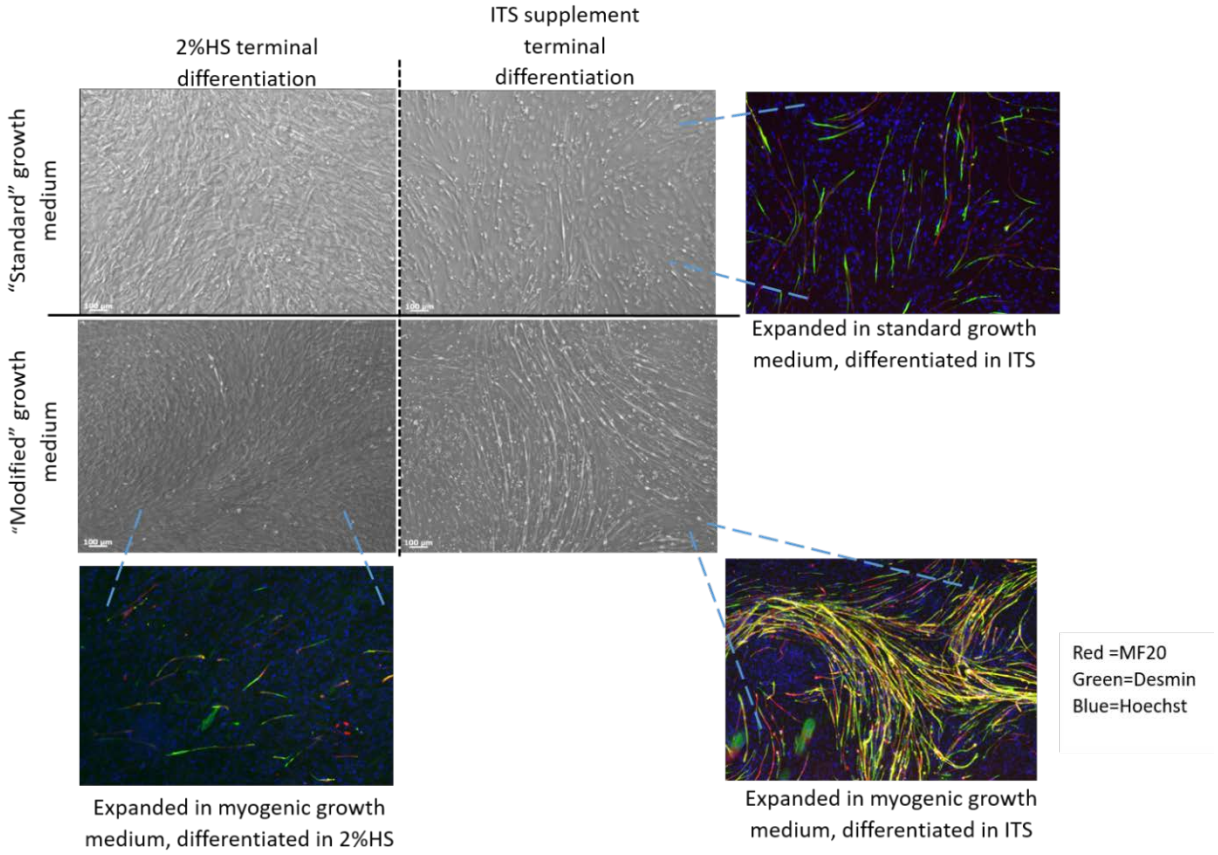



**Figure 4.6: Optimization of culture conditions for the selective expansion and terminal specification of myogenic progenitors**

Brightfield images depict four culture conditions for pairwise testing of two growth media, followed by two differentiation conditions. IF staining shows that more long, multinucleated Mf20/desmin positive cells arise in culture conditions of FGF2-containing growth medium, followed by serum-free ITS-supplemented differentiation medium. We optimized the conditions using cells that were differentiated up to 30 days using the protocol by Borchin et al.

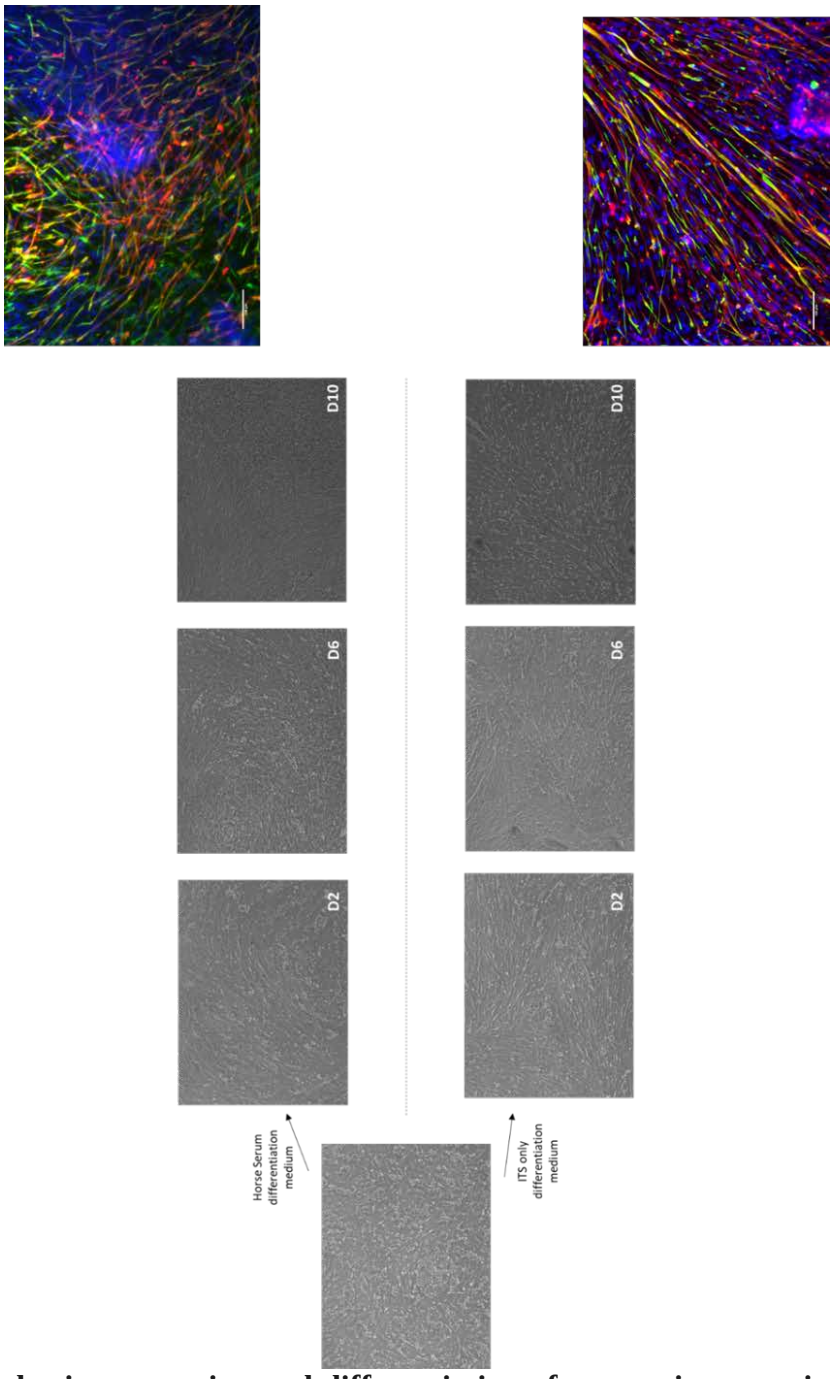
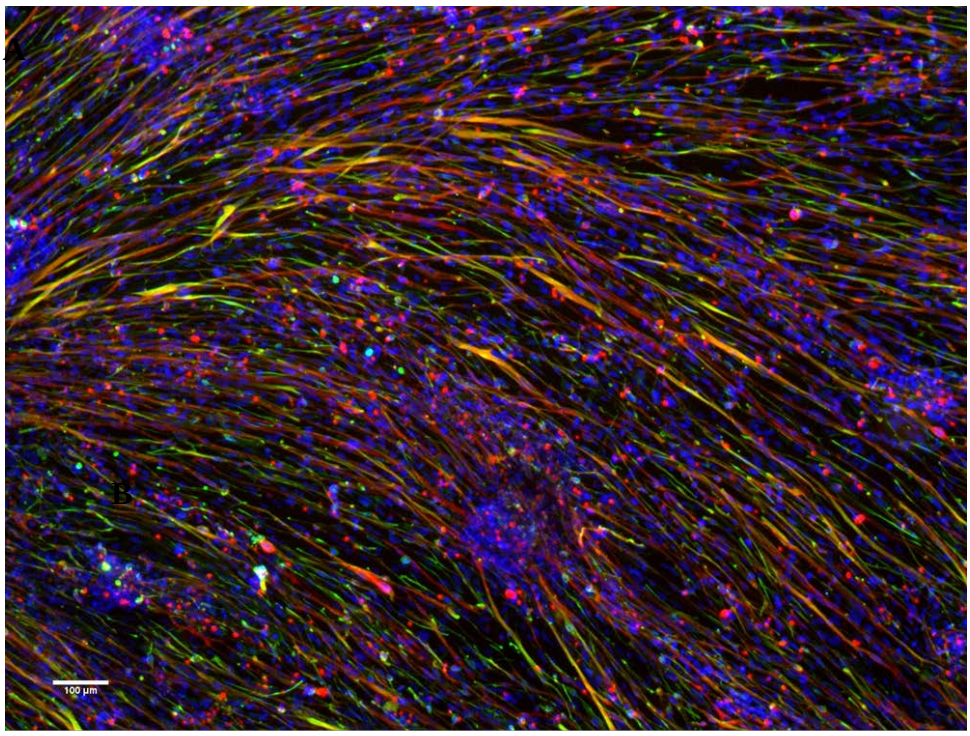



**Figure 4.7: Selective expansion and differentiation of myogenic progenitors derived for 30 days using the protocol by Chal et al.**

We passaged into FGF2-supplemented growth medium cells that were specified to the myogenic lineage for 30 days using the protocol by Chal et al. We again tested 2% horse serum differentiation medium and ITS-supplemented differentiation media and observed more multinucleated, aligned Mf20/desmin positive cells in the ITS_supplemented differentiation medium.




**B**

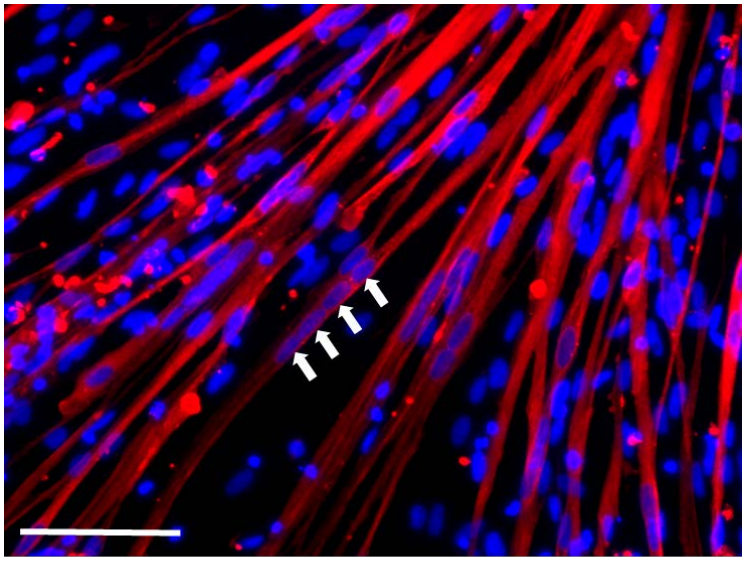



**Figure 4.8: Optimized culture conditions yield purified population of multinucleated myogenic progenitors**

A. 5x immunofluorescent staining for Mf20 and desmin shows a relatively homogeneous population of long multinucleated cells after passaging, selective expansion, and terminal differentiation according to our modified protocol, after 30 days of initial differentiation according to the procotol by Chal et al. Scale bar = 100um

B. 20x image depicts numerous multinucleated cells. Scale bar = 100um

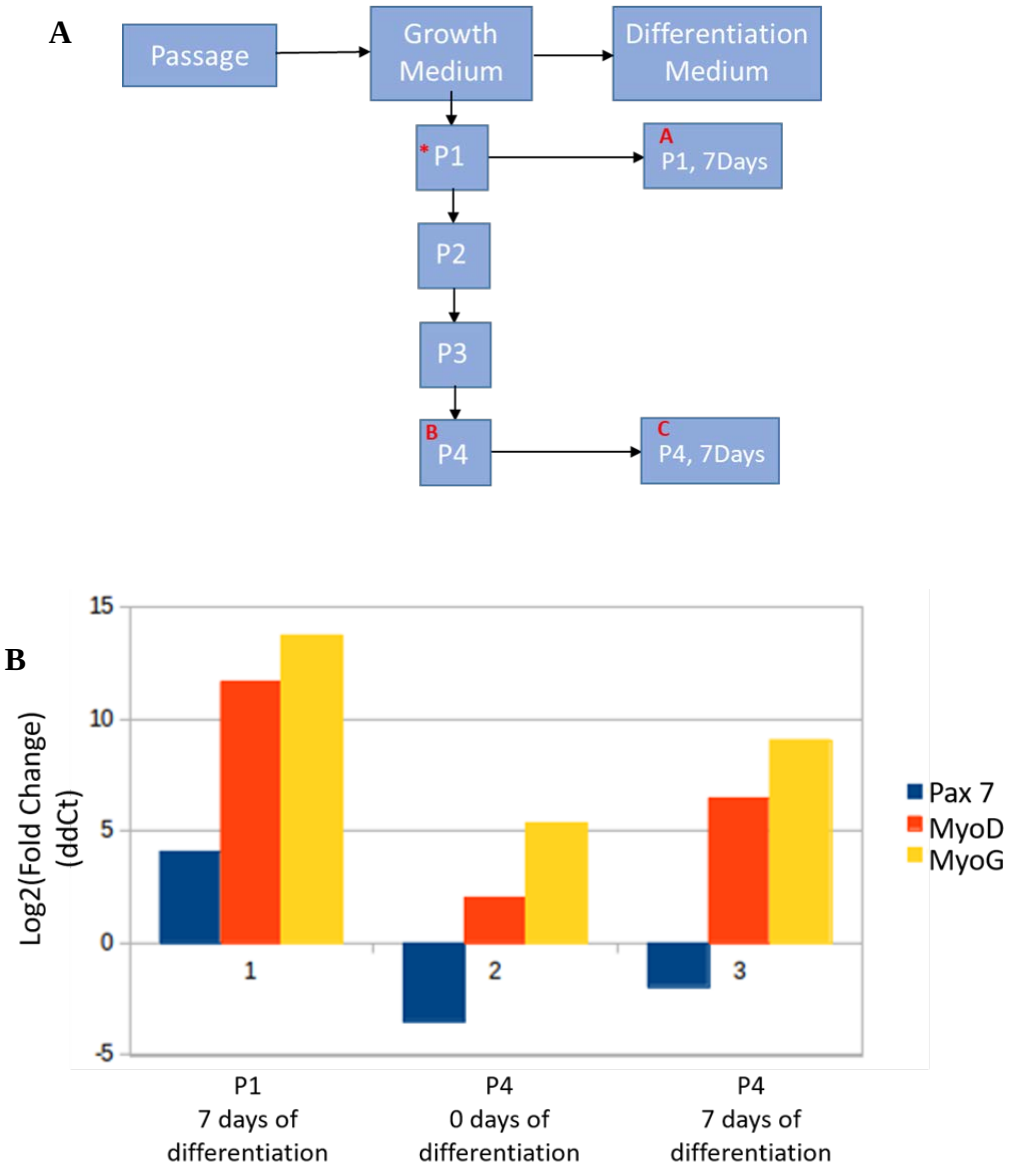



**Figure 4.9: Passage number-dependent gene expression of MyoD, MyoG, and Pax7**

A. Schematic for testing the retention of myogenic nature with serial passaging. After 30 days of initial differentiation, we compared the gene expression of MyoD, MyoG, and Pax7 in myogenic progenitors upon serial passaging up to four passages. We compared their ability to differentiate at early passage (P1) and at late passage (P4). Gene expression is normalized to undifferentiated P1.

B. Gene expression shows that cells retain some ability to differentiate at P4, as evidenced by the increase of MyoD and MyoG upon differentiation. However, Pax7 expression is attenuated at higher passage.

**Figure 4.10: Immunofluorescent characterization of early- and late- specified myogenic progenitor populations**

The early population is uninuclear, Myf5+/MyoG-/Mf20-/dystrophin-/Pax7-, while the late population is comprised of a mix of multinucleated My5+/MyoG+/Mf20+/Dystrophin+ cells, and uninucleated Pax7+ cells.

E. MyoG/Myf5

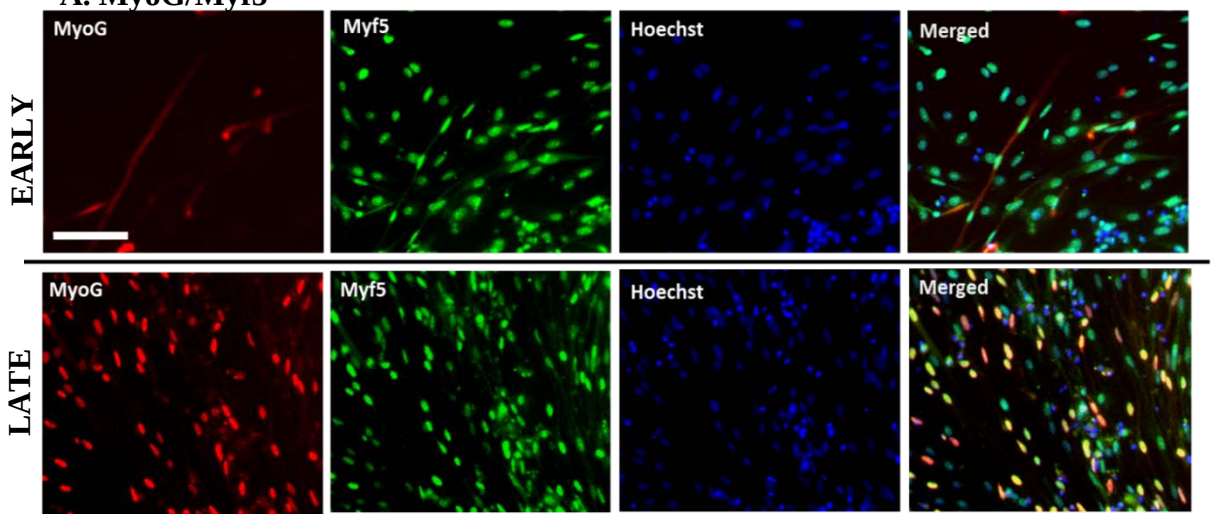
F. Mf20/Desmin

G. Pax7/Desmin

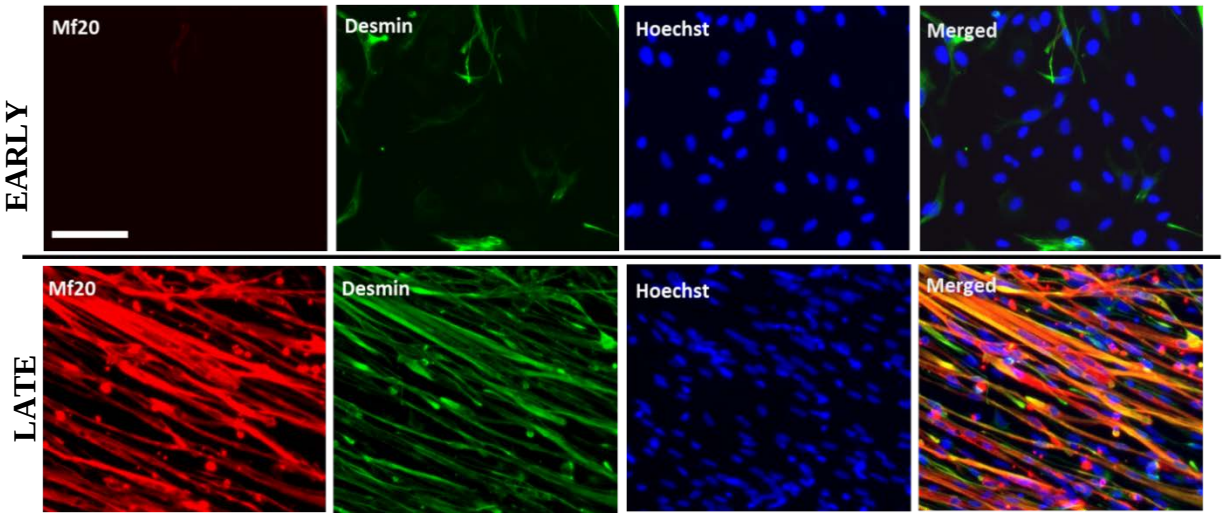
H. Dystrophin

Scale bars=100um

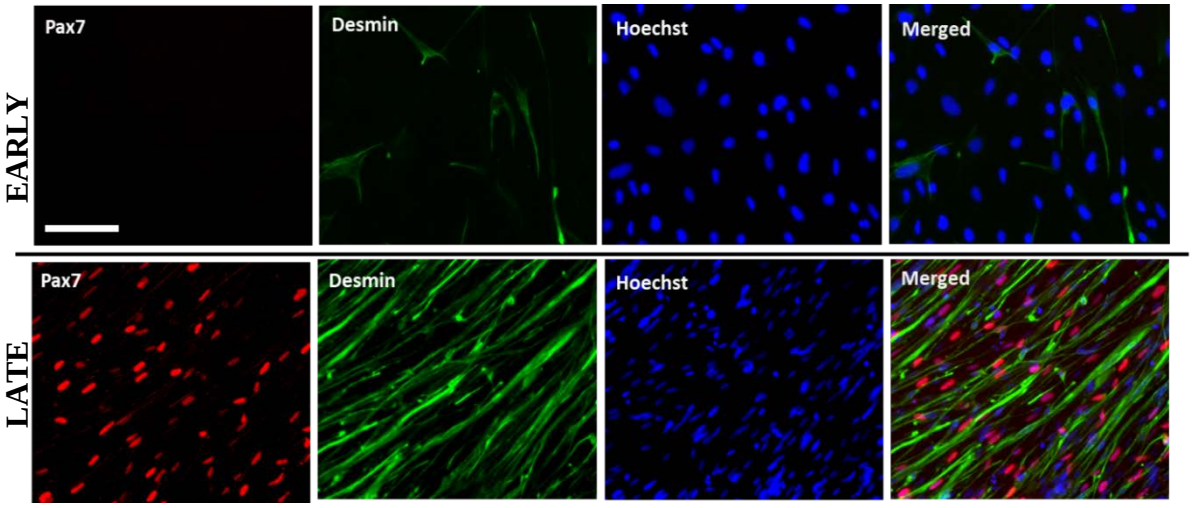
**A. MyoG/Myf5**




**B. Mf20/Desmin**



**C. Pax7/Desmin**




**D. Dystrophin**

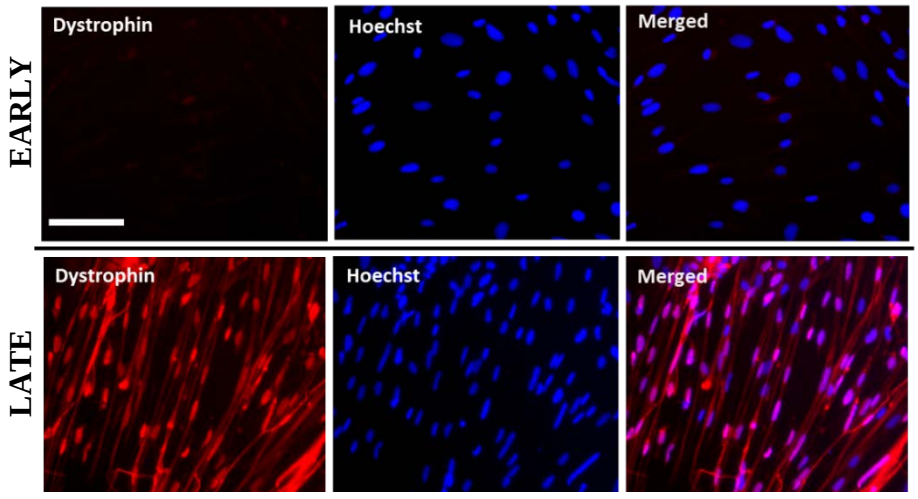
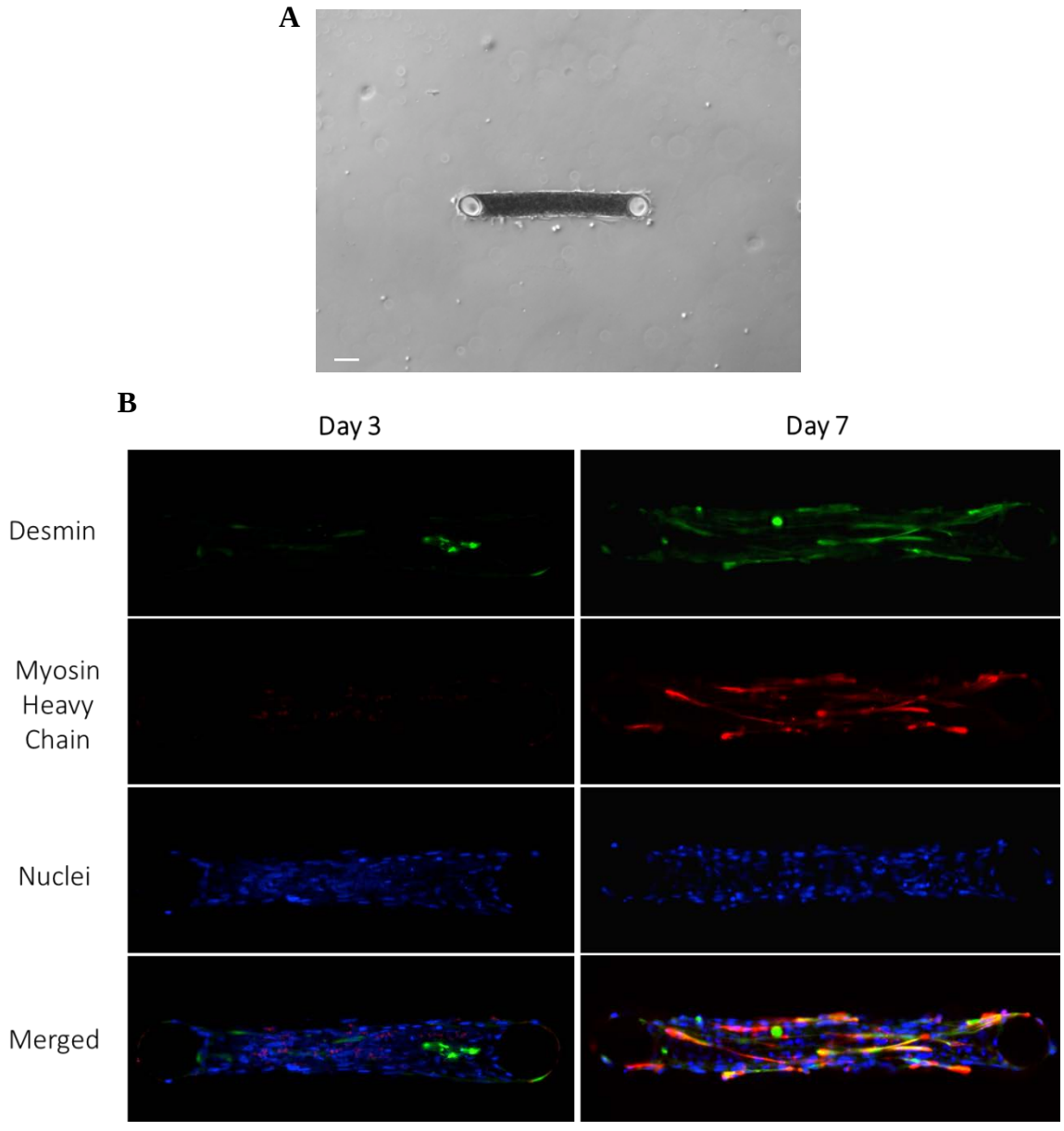


**Figure 4.11: Formation of hiPSC-derived myogenic microtissues on an in vitro skeletal muscle-on-chip platform**
A. Brightfield image of formed microtissue. Pillars at either end anchor the microtissue and allow for alignment and compaction, as well as providing support for the 3D tissue.
B. Maturation of tissue from 3 to 7 days of differentiation in the chip device, as evidenced by increased Mf20 and desmin positive cells at day 7.


A
B
Day 3
Day 7
Desmin
Myosin Heavy Chain
Nuclei
Merged

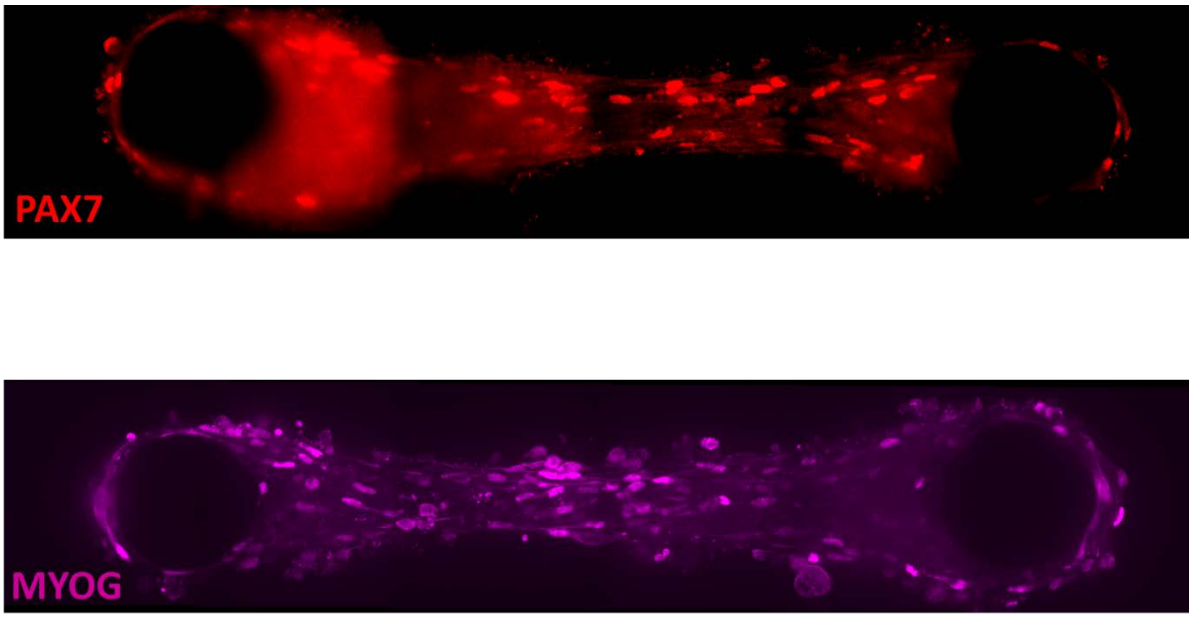



**Figure 4.12: Formation of hiPSC-derived myogenic microtissues on an in vitro skeletal muscle-on-chip platform**
PAX7 (above) and MyoG (below) positive cells within the maturing microtissue.

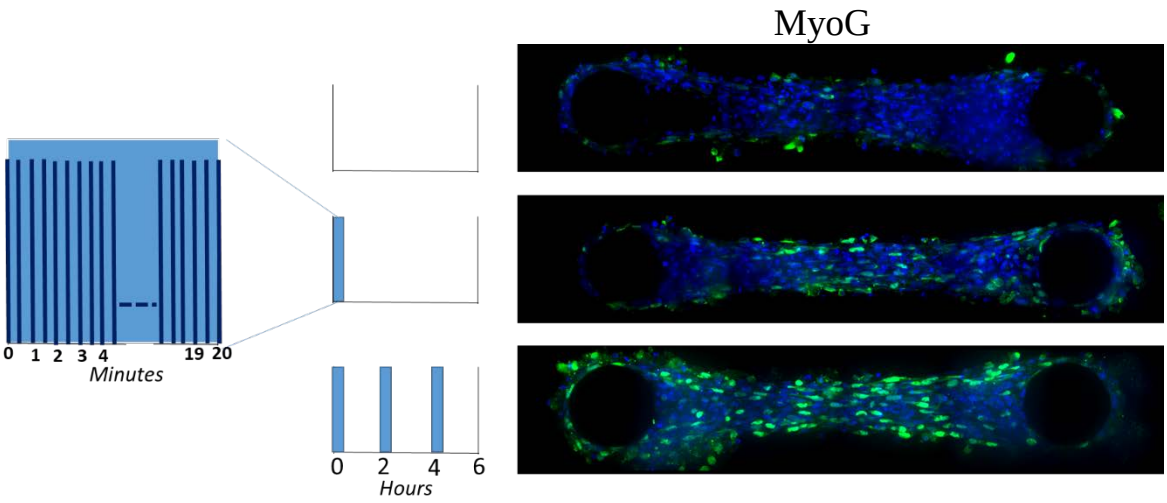



**Figure 4.13 Application of periodic, passive mechanical loading hastens maturation of hiPSC derived myogenic microtissue in a dose-dependent fashion**
We observed an increase in the number of MyoG positive cells in the microtissue with respect to increasing frequency of passive, mechanical loading from once every six hours to three times every six hours. Each stimulation consisted of 40 pulses delivered over 20 minutes.

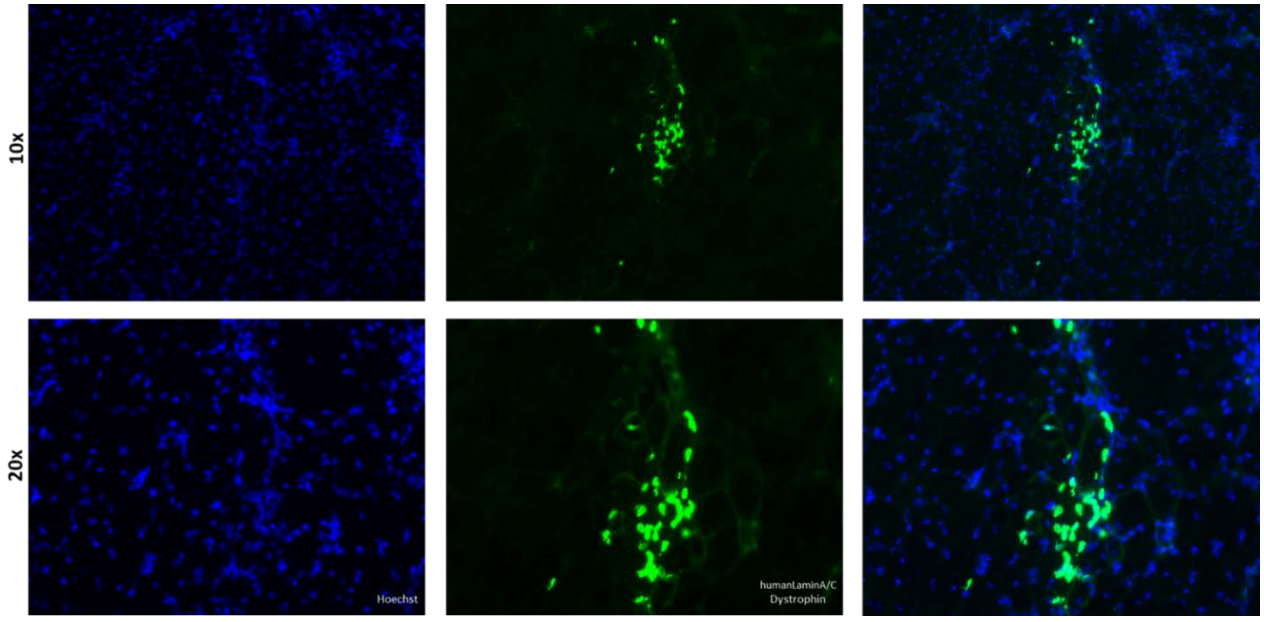



**Figure 4.14: Transplanted hiPSC-derived myogenic progenitors give rise to dystrophin positive fibers in murine model of muscular dystrophy**
Dystrophin positive fibers co-localize with human lamin A/C positive cells 28 days post-transplantation.

## Methods

### Pluripotent iPSC culture

iPSC lines were derived from peripheral blood cells. Pluripotent hiPSC colonies were maintained in feeder free conditions. Plates were coated with Matrigel (Corning) according to manufacturer's instructions and mTeSR1 (STEMCELL Tech) maintenance media changed daily was used to grow colonies. Cells were detached for replating as colonies using ReLeSR according to manufacturer's instructions. Cells were seeded as single cell by detaching with Versene (0.5 mM EDTA solution) and replated in mTeSR for ~24 hours in the presence of 2uM TZV before switching to differentiation medium.

### Myogenic differentiation

Cells seeded as single cells on Matrigel coated plates at a seeding density of 20,000-30,000 cells per square centimeter. Five differentiation media conditions were used over the course of 35 days of myogenic differentiation. Days 1-3: DMEM/F12 was mixed with 1% NEAA, 1% Glutamax, 1% pen/strep, 1% ITS supplement, CHIR99021 (3uM), LDN-193189 (0.5uM). Days 4-6: DMEM/F12 was mixed with 1% NEAA, 1%Glutamax, 1% pen/strep, 1%ITS supplement, CHIR99021 (3uM), LDN-193189 (0.5uM), FGF (20ng/mL). Days 7-8: DMEM/F12, 15% KOSR (v/v), 1% NEAA, 1% Glutamax, 1% pen/strep, HGF (10 ng/mL), IGF (2ng/mL), FGF (20ng/mL), LDN (0.5uM), beta mercaptoethanol (0.1mM). Days 9-12: DMEM/F12, 15% KOSR, 1% NEAA, 1%Glutamax, 1% pen/strep, IGF (2ng/mL), beta mercaptoethanol (0.1mM). Days 13-35: DMEM/F12, 15% KOSR, 1% NEAA, 1%Glutamax, 1% pen/strep, HGF (10 ng/mL), IGF (2ng/mL), beta

mercaptoethanol (0.1mM). Media changed daily for the first 12 days, followed by half media change every day for the remainder of the differentiation

Passaging, expanding, and terminal differentiation

To passage differentiated cells pretreat wells with 2uM TZV for about 2 hours before passaging. Detach cells with TrypLE Express by incubating cells for five minutes at 37C. Replate cells in modified growth medium: DMEM/F12+20%KOSR (v/v)+1% NEAA, 1% Glutamax, 1% pen/strep+2.5ng/mL FGF. One well of differentiated cells should be passaged into 3 wells. When passaged cells reach 70-80% confluence, switch to differentiation medium: DMEM/F12+1% NEAA, 1%Glutamax, 1% pen/strep, 1%ITS supplement.

Immunofluorescent staining

Immunofluorescent staining was performed using the following primary antibodies: PAX3, PAX7, MF20 (1:200; Developmental Studies Hybridoma Bank), MYF5 (1:200; Santa Cruz Biotechnology), desmin (1:200; Abcam), and human lamin A/C (1:50; Vector Laboratories). The following secondary antibodies were used: goat anti-rat Alexa 546 (1:200; Life Technologies), goat anti-mouse Alexa488 (1:250; Life Technologies), and goat anti-rabbit Alexa 546 (1:200; Life Technologies). For immunofluorescent staining of cells grown on tissue culture plates, cells were fixed in 4% PFA for 10 min at room temperature. Immediately before staining, the cells were permeabilized with 0.1% (v/v) Triton X-100 and blocked with 3% (w/v) BSA for 30 mins. Cells were stained with primary antibodies diluted in 1% BSA overnight at 4°C, washed 3 times with PBS, and incubated with secondary antibodies for 1 hr at room temperature. The nuclei were stained with

Hoechst 33342 (2 μg/ml; Life Technologies) for 5 mins at room temperature. For immunofluorescent staining of tibialis anterior (TA) muscles, samples were first embedded in optimal temperature cutting compound (OCT) for cryosectioning and sections (having around 20 μm thickness) were fixed with 4% PFA for 10 mins at room temperature. Next, sections were permeabilized with 0.3% Triton X-100, blocked with 3% BSA for 1 hr at room temperature, and stained with human lamin A/C. Imaging was performed using a fluorescence microscope (Carl Zeiss; Axio Observer A1)

RT-qPCR

hiPSCs cultured for various amount of time according to the myogenic differentiation protocol were examined for changes in the gene expression as a function of culture time. RNA was extracted from cell cultures using TRIzol according to the manufacturer's instructions. For each sample, 1 μg of RNA was reverse-transcribed to complementary DNA (cDNA) using an iScript cDNA synthesis kit (Bio-Rad, catalog no. 170-8891). Real-time PCR reactions were run on ABI Prism 7700 Real-time PCR Cycler (Applied Biosystems).For qPCR analysis of selective genes, SYBR Select Master Mix (Life Technologies, catalog no. 4472908) was mixed with various primers (Pax7, Myf5, MyoD, MyoG). The expression of each target gene was normalized to that of corresponding 18S, a housekeeping gene. The expression levels were normalized to that of undifferentiated, pluripotent hPSCs and presented as log2(fold change).

Cell transplantation

24 hours prior to cell transplantation, the TA muscles of 2-month-old immune-deficient NOD.CB17-Prkdc$^{scid}$/J mice (NOD/SCID mice) were injured using cardiotoxin. The mice

were first anesthetized by injecting ketamine (100 mg/kg) and xylazine (10 mg/kg) intraperitoneally and then injected with 30 µL of 10 mM cardiotoxin into the TA muscle. Early- and late- extent of differentiation cell populations were suspended in physiological saline solution and injected into the TA muscles. Each muscle was injected with about 300,000 cells suspended in 40 uL of physiologial saline solution. All animal studies were carried out with the approval of Institutional Animal Care and Use Committee (IACUC) of the University of California, San Diego. Two and four days following transplantation, muscles were harvested and the *in vivo* viability of the donor cells was assessed histologically.

## References


1. Kim JH, Ko IK, Atala A, Yoo JJ. Progressive Muscle Cell Delivery as a Solution for Volumetric Muscle Defect Repair. *Scientific Reports*. 2016;6:38754. doi:10.1038/srep38754.

2. Sienkiewicz D, Kulak W, Okurowska-Zawada B, Paszko-Patej G, Kawnik K. Duchenne muscular dystrophy: current cell therapies. *Therapeutic Advances in Neurological Disorders*. 2015;8(4):166-177. doi:10.1177/1756285615586123.

3. Hwang Y, Suk S, Lin S, et al. Directed *In Vitro* Myogenesis of Human Embryonic Stem Cells and Their *In Vivo* Engraftment. Emanueli C, ed. *PLoS ONE*. 2013;8(8):e72023. doi:10.1371/journal.pone.0072023.

4. Cosgrove BD, Gilbert PM, Porpiglia E, et al. Rejuvenation of the aged muscle stem cell population restores strength to injured aged muscles. *Nature medicine*. 2014;20(3):255-264. doi:10.1038/nm.3464.

5. Baraniak PR, McDevitt TC. Stem cell paracrine actions and tissue regeneration. *Regenerative medicine*. 2010;5(1):121-143. doi:10.2217/rme.09.74.

6. Trounson A, Thakar RG, Lomax G, Gibbons D. Clinical trials for stem cell therapies. *BMC Medicine*. 2011;9:52. doi:10.1186/1741-7015-9-52.

7. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006 Aug 25;126(4):663-76. Epub 2006 Aug 10.

8. Madden L, Juhas M, Kraus WE, Truskey GA, Bursac N. Bioengineered human myobundles mimic clinical responses of skeletal muscle to drugs. Wagers AJ, ed. *eLife*. 2015;4:e04885. doi:10.7554/eLife.04885.

9. Bellin M, Marchetto MC, Gage FH, Mummery CL. Induced pluripotent stem cells: the new patient? *Nat Rev Mol Cell Biol.* 2012 Nov;13(11):713-26. doi: 10.1038/nrm3448.

10. Kodaka Y, Rabu G, Asakura A. Skeletal Muscle Cell Induction from Pluripotent Stem Cells. *Stem Cells International*. 2017;2017:1376151. doi:10.1155/2017/1376151.

11. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell. 1987 Dec 24;51(6):987-1000.

12. Abujarour R, Bennett M, Valamehr B, et al. Myogenic Differentiation of Muscular Dystrophy-Specific Induced Pluripotent Stem Cells for Use in Drug Discovery. *Stem Cells Translational Medicine*. 2014;3(2):149-160. doi:10.5966/sctm.2013-0095.

13. Young CS, Hicks MR, Ermolova NV, et al. A single CRISPR-Cas9 deletion strategy that targets the majority of DMD patients restores dystrophin function in hiPSC-derived muscle cells. *Cell stem cell*. 2016;18(4):533-540. doi:10.1016/j.stem.2016.01.021.

14. Goudenege S, Lebel C, Huot NB, et al. Myoblasts Derived From Normal hESCs and Dystrophic hiPSCs Efficiently Fuse With Existing Muscle Fibers Following Transplantation. *Molecular Therapy*. 2012;20(11):2153-2167. doi:10.1038/mt.2012.188.

15. Pettinato G, Wen X, Zhang N. Formation of Well-defined Embryoid Bodies from Dissociated Human Induced Pluripotent Stem Cells using Microfabricated Cell-repellent Microwell Arrays. *Scientific Reports*. 2014;7402 doi:10.1038/srep07402.

16. Borchin B, Chen J, Barberi T. Derivation and FACS-Mediated Purification of PAX3+/PAX7+ Skeletal Muscle Precursors from Human Pluripotent Stem Cells.*Stem Cell Reports*. 2013;1(6):620-631. doi:10.1016/j.stemcr.2013.10.007.

17. Chal J, Al Tanoury Z, Hestin M, Gobert B, Aivio S, Hick A, Cherrier T, Nesmith AP, Parker KK, Pourquié O. Generation of human muscle fibers and satellite-like cells from human pluripotent stem cells *in vitro* Nat Protoc. 2016 Oct;11(10):1833-50. doi: 10.1038/nprot.2016.110.

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

19.Hall JK, Banks GB, Chamberlain JS, Olwin BB. Prevention of Muscle Aging by Myofiber-Associated Satellite Cell Transplantation. *Science translational medicine*. 2010;2(57):57ra83. doi:10.1126/scitranslmed.3001081.

20. Dumont NA, Wang YX, Rudnicki MA. Intrinsic and extrinsic mechanisms regulating satellite cell function. *Development (Cambridge, England)*. 2015;142(9):1572-1581. doi:10.1242/dev.114223.

21. Bentzinger CF, Wang YX, Rudnicki MA. Building Muscle: Molecular Regulation of Myogenesis. *Cold Spring Harbor Perspectives in Biology*. 2012;4(2):a008342. doi:10.1101/cshperspect.a008342.

22. Kuang S, Gillespie MA, Rudnicki MA. Niche regulation of muscle satellite cell self-renewal and differentiation. Cell Stem Cell. 2008 Jan 10;2(1):22-31. doi: 10.1016/j.stem.2007.12.012.

23. Rudnicki MA, Le Grand F, McKinnell I, Kuang S. The molecular regulation of muscle stem cell function. Cold Spring Harb Symp Quant Biol. 2008;73:323-31. doi: 10.1101/sqb.2008.73.064.

24. Osses N, Brandan E. ECM is required for skeletal muscle differentiation independently of muscle regulatory factor expression. Am J Physiol Cell Physiol. 2002 Feb;282(2):C383-94.

25. Chaturvedi V, Dye DE, Kinnear BF, van Kuppevelt TH, Grounds MD, Coombe DR. Interactions between Skeletal Muscle Myoblasts and their Extracellular Matrix Revealed by a Serum Free Culture System. Engler AJ, ed. *PLoS ONE*. 2015;10(6):e0127675. doi:10.1371/journal.pone.0127675.

26. Thomas K, Engler AJ, Meyer GA. Extracellular Matrix Regulation in the Muscle Satellite Cell Niche. *Connective tissue research*. 2015;56(1):1-8. doi:10.3109/03008207.2014.947369.

27. Mukund K, Mathewson M, Minamoto V, Ward SR, Subramaniam S, Lieber RL. Systems Analysis of Transcriptional Data Provides Insights Into Muscle's Biological Response to Botulinum Toxin. *Muscle & nerve*. 2014;50(5):744-758. doi:10.1002/mus.24211.

28. Aguilar CA, Shcherbina A, Ricke DO, et al. *In vivo* Monitoring of Transcriptional Dynamics After Lower-Limb Muscle Injury Enables Quantitative Classification of Healing. *Scientific Reports*. 2015;5:13885. doi:10.1038/srep13885.

29. Kim JY, Park Y-K, Lee K-P, et al. Genome-wide profiling of the microRNA-mRNA regulatory network in skeletal muscle with aging. *Aging (Albany NY)*. 2014;6(7):524-544.

30. Palstra AP, Beltran S, Burgerhout E, et al. Deep RNA Sequencing of the Skeletal Muscle Transcriptome in Swimming Fish. Rawls JF, ed. *PLoS ONE*. 2013;8(1):e53171. doi:10.1371/journal.pone.0053171.

31. Ayuso M, Fernández A, Núñez Y, et al. Comparative Analysis of Muscle Transcriptome between Pig Genotypes Identifies Genes and Regulatory Mechanisms Associated to Growth, Fatness and Metabolism. PENA i SUBIRÀ RN, ed. *PLoS ONE*. 2015;10(12):e0145162. doi:10.1371/journal.pone.0145162.

32. Tripathi AK, Patel AK, Shah RK, Patel AB, Shah TM, Bhatt VD, Joshi CG. Transcriptomic dissection of myogenic differentiation signature in caprine by RNA-Seq. Mech Dev. 2014 May;132:79-92. doi: 10.1016/j.mod.2014.01.001.

33. Jingting S, Qin X, Yanju S, Ming Z, Yunjie T, Gaige J, Zhongwei S, Jianmin Z. Oxidative and glycolytic skeletal muscles show marked differences in gene expression profile in Chinese Qingyuan partridge chickens. PLoS One. 2017 Aug 16;12(8):e0183118. doi: 10.1371/journal.pone.0183118.

34. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*. 2012;7(3):562-578. doi:10.1038/nprot.2012.016.

35. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

36. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi:10.1093/bioinformatics/btw354.

37. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017 Apr;14(4):417-419. doi: 10.1038/nmeth.4197.

38. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. 2015;4:1521. doi:10.12688/f1000research.7563.2.

39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8.

40. Murphy MB, Moncivais K, Caplan AI. Mesenchymal stem cells: environmentally responsive therapeutics for regenerative medicine. Experimental & Molecular Medicine. 2013;45(11):e54-. doi:10.1038/emm.2013.94.

41.Wayne AS, Baird K, Egeler RM. Pediatric Clinics of North America: Hematopoietic Stem Cell Transplantation Stem Cell Transplantation for Leukemia. Pediatric clinics of North America. 2010;57(1):1-25. doi:10.1016/j.pcl.2009.11.005.

42. Hügle T, Daikeler T. Stem cell transplantation for autoimmune diseases. Haematologica. 2010;95(2):185-188. doi:10.3324/haematol.2009.017038.

43. Kochenderfer JN, Dudley ME, Feldman SA, et al. B-cell depletion and remissions of malignancy along with cytokine-associated toxicity in a clinical trial of anti-CD19 chimeric-antigen-receptor–transduced T cells. Blood. 2012;119(12):2709-2720. doi:10.1182/blood-2011-10-384388.

44. Moat SJ, Bradley DM, Salmon R, Clarke A, Hartley L. Newborn bloodspot screening for Duchenne Muscular Dystrophy: 21 years experience in Wales (UK). European Journal of Human Genetics. 2013;21(10):1049-1053. doi:10.1038/ejhg.2012.301.

45. Meng J, Bencze M, Asfahani R, Muntoni F, Morgan JE. The effect of the muscle environment on the regenerative capacity of human skeletal muscle stem cells. Skeletal Muscle. 2015;5:11. doi:10.1186/s13395-015-0036-8

46. Kabra H, Hwang Y, Lim HL, Kar M, Arya G, Varghese S. Biomimetic Material-Assisted Delivery of Human Embryonic Stem Cell Derivatives for Enhanced In Vivo Survival and Engraftment. ACS biomaterials science & engineering. 2015;1(1):7-12. doi:10.1021/ab500021a.

47. Hwang Y, Suk S, Shih Y-RV, et al. WNT3A promotes myogenesis of human embryonic stem cells and enhances in vivo engraftment. Scientific Reports. 2014;4:5916. doi:10.1038/srep05916.

48. Sampaolesi M, Blot S, D'Antona G, Granger N, Tonlorenzi R, Innocenzi A, Mognol P, Thibaud JL, Galvez BG, Barthélémy I, Perani L, Mantero S, Guttinger M, Pansarasa O, Rinaldi C, Cusella De Angelis MG, Torrente Y, Bordignon C, Bottinelli R, Cossu G.Mesoangioblast stem cells ameliorate muscle function in dystrophic dogs. Nature. 2006 Nov 30;444(7119):574-9

49. Cossu G, Previtali SC, Napolitano S, et al. Intra-arterial transplantation of HLA-matched donor mesoangioblasts in Duchenne muscular dystrophy. EMBO Molecular Medicine. 2015;7(12):1513-1528. doi:10.15252/emmm.201505636.

50. Sacco A, Doyonnas R, Kraft P, Vitorovic S, Blau HM. Self-renewal and expansion of single transplanted muscle stem cells. Nature. 2008;456(7221):502-506. doi:10.1038/nature07384.

# Chapter 5: Conclusions and future directions

In this dissertation, we have addressed several challenges associated with the myogenic differentiation of human induced pluripotent stem cells. In the first two aims, we took a step towards understanding the mechanisms behind robust versus attenuated in vitro myogenesis. We generated and analyzed longitudinal transcriptomic data from three hiPSC lines (L, S, and T) over the course of thirty days of myogenic induction to uncover causative mechanisms of myogenesis in a time-dependent fashion. IF staining of the three cell lines revealed that one cell line (L) had earlier expression of several markers of myogenic specification, including MyoG, myosin heavy chain, and Pax7, compared to the other two lines (S,T) in the latter third of differentiation. However, clustering of the gene expression data with respect to time points showed a divergence between the L line compared to the S and T lines starting around day 6, and this discrepancy increased with time of differentiation. From this, by investigating the cell-line and time-dependent expression of several key transcription factors, we showed that a crucial divergence in lineage specification occurred at the outset of induction, at the point of commitment to an initial germ layer. We then implicated several major chromatin modifying complexes in the regulation of exit from pluripotency, and further found cell-line dependent expression of several changeable components of the complexes. This observation supports several reports in the literature that tie context-specific composition of the complexes with differential targeting to genomic sites. We also noted cell-line dependent expression of numerous components of major signaling pathways, especially in the early part of differentiation. Various studies of in vitro hiPSC differentiation to several target cell

lineages- including hepatic, kidney, cardiac, pancreatic, and myogenic- describe variable temporal activation/inhibition and concentration modulation of major signaling pathways to achieve initial germ layer specification. This indicates, first that differentiation is extremely context specific, and second that there may be downstream transcriptional regulators that serve to integrate a host of signals and, through cooperation with chromatin remodelers/modifiers, affect the gene transcription that leads to lineage specification. Since Wnt/B-catenin activation is known to guide paraxial streak formation in vivo, and since treatment with CHIR, a GSK3B inhibitor, is the foundation of several protocols for the in vitro generation of mesendoderm progenitors from hPSCs, we focused our genetic perturbation experiments on transcriptional cofactors/ repressors of B-catenin that had cell-line dependent expression at the outset of differentiation. SiRNA mediated knockdown of the B-catenin repressor TLE6 and the activator LEF1 led to enhanced and attenuated paraxial mesoderm, respectively. We showed that modulation of the transcriptional activity of B-catenin could lead to more robust paraxial mesoderm specification, and more efficient downstream myogenesis, which could be of utility for clinical applications of myogenic progenitors.

There are relatively few longitudinal transcriptomic studies that track the differentiation of multiple hiPSC lines; to our knowledge, this is the first for myogenic specification. Progress in the transgene-free, directed myogenic differentiation of hPSCs has lagged behind the progress of differentiation towards other lineages. Recent protocols have used mouse and human embryology as a guide to refine the small molecule and growth factors cues needed to direct robust myogenic differentiation, with some success.

However, these approaches miss initial germ layer specification since they use the tailbud as the baseline for somite development. By contrast, pluripotent hiPSCs are closer to the pre-gastrulation epiblast in terms of their gene expression and epigenetic profiles. In addition, though in vitro differentiation may be analogous to in vivo development in some respects, there are numerous crucial differences particularly in cell-cell interactions including paracrine effects, signaling ligand gradients, monolayer culture versus 3D growth, and timing. Therefore, our dataset can provide valuable insight to in vitro specific processes.

Time-dependent gene expression profiling that compares multiple hiPSC lines in crucial to the interpretation of the data. In vitro myogenic differentiation is known to be heterogeneous both in terms of efficiency of differentiation between cell lines, as well as in terms of the variety of off-target/contaminating cell lineages that may arise over the course of even a robust differentiation. We used the time-varying expression profiles across all three cell lines to group together genes with significant functional enrichment. In addition to segregating heterogeneous processes to some extent, these gene modules also served to indicate the function of numerous annotated and unannotated transcripts, including noncoding RNAs and transcription factors in the context of differentiation; chief among these were components that may serve to target chromatin modifying complexes and cofactors of beta catenin that might integrate numerous signals. Using multiple hiPSC lines also helped to separate out cell line specific idiosyncratic gene expression.

Indeed, several of our observations may warrant further detailed study. First, although we have implicated the components of the chromatin modifying complexes, we

have not shown definitively that they indeed play a role in the promotion or blunting of paraxial mesoderm specification. Furthermore, it would be interesting to know whether any of these components interact with B-catenin and/or the core transcriptional regulators of pluripotency to effect chromatin compaction or opening. To build on our gene knockdown experiments, such studies would optimally include gene expression profiling coupled with assays of chromatin structure, such as ATAC-seq or HiC to tie gene expression to chromatin topology, in the context of modified B-catenin signaling. If a transcription factor of interest is identified, ChiP or ChiP-Seq would show specific transcription factor binding. One step further, CHiP-CHiP could be used to look for co-localization of a transcription factor and a particular chromatin modifying complex. A previous study has detailed the effect of varying concentrations of agonists of several signaling pathways, including Wnt, BMP, and FGF on definitive endoderm commitment. A similar approach, coupled with study of B-catenin cofactors and key Yamanaka factors, as described above, could shed light on how the B-catenin gene regulatory network acts to integrate multiple signals to regulate gene expression.

Second, our study has also identified unannotated transcripts that may play a role at various stages of myogenic specification, through their clustering with known genes. In particular, lncRNAs have been shown to participate at several levels of transcriptional regulation, including as a sponge for miRNAs, targeting chromatin modifying complexes for enhanced function, inhibiting chromatin modifier function by direct binding, and finally physically linking sections of chromatin to act as a scaffold in trans-regulation. Still, the majority of lncRNAs remain unannotated, and some even hypothesize that many of these

transcripts may represent transcriptional noise. However, in our study, we can focus on those transcripts that are more likely to represent regulatory elements if they are expressed in a similar time course in multiple cell lines. An interesting first step could be computational. Unsupervised clustering that combines information from our dataset- gene modules with genes of known function including known lncRNAs- with information culled from databases could further narrow the list of promising candidate lncRNAs that may play regulatory roles. Information from databases includes lncRNA secondary structure and sequence, potential expression in other transcriptomic datasets, possible tissue-specific expression, and the genomic neighborhood of the transcript of interest. If several unannotated lncRNAs of interest are identified, they could be studied further via knockdown with antisense oligonucleotides and chromatin structure/ TF binding assays as described above.

Finally, although we have focused heavily on germ layer specification, we have not delved into terminal myogenesis in much detail. Part of the reason for this is that with increasing time of differentiation, the cell population increased in heterogeneity so it was more difficult to identify groups of genes that played a role in supporting myogenesis. (The core, well-known myogenic regulatory factors could readily be grouped). In particular, it would be interesting to look further at the role that ancillary cell populations such as neurons and fibroblasts may have on supporting in vitro myogenesis. Furthermore, since we observe the appearance of Pax7+ cells, it would be interesting to study further the gene expression that might accompany the creation of the satellite cell niche, such as secretion of ECM proteins or paracrine growth factors, cytokines that might support the culture of

quiescent satellite cells. One way to study the latter point in more detail would be to use the modified protocol we developed in aim 3. This allows for the purification and terminal differentiation of myogenic progenitors, and we also observe the appearance of Pax7 positive cells following MyoG expression. Candidate genes from the RNA seq data could then be tested in this more controlled context. Insight on the spontaneous creation of a niche that supports the in vitro expansion of satellite cells could be of great utility, since their in vitro expansion and maintenance is extremely limited by differentiation.

A key challenge in carrying out the above studies is the heterogeneity of differentiation. Assays of chromatin structure would have to be carried out on a relatively pure cell population in order to have meaning. In the case of bulk RNAseq, as in the present study, we are able to assess the functional enrichment of gene groups clustered according to their temporal expression patterns across multiple cell lines. However, this approach may not work for studies for chromatin structure since far fewer patterns of chromatin topology have been linked to a particular biological function- not to mention the added cost, time, and complexity introduced in a longitudinal study across multiple cell lines. One possible solution could be the use of a reporter line that would fluoresce in response to the expression of a transcription factor of interest. Several hESC and hiPSC lines in the literature report brachyury expression; however, brachyury expression is not specific to paraxial mesoderm, so markers in addition to brachyury would have to be used. Our current study could also aid in the creation of specialized reporter lines that are specific for paraxial mesoderm induction and downstream myogenic specification. Knowledge of the relative temporal expression patterns of key genes would help to delineate paraxial mesoderm

progenitors from other primitive streak derivatives. In addition to multiple markers for selection purposes, expression of these markers could help gauge the progress of myogenic induction; this insight would likely save time and help streamline the optimization of detailed, cell-line specific culture conditions.

In the third aim, we modified an existing protocol for the purification, expansion, and terminal differentiation of myogenic progenitors. Furthermore, we validated the ability of these progenitors to form functional in vitro microtissues in a skeletal muscle on a chip platform, as well as their in vivo function in a murine model of muscular dystrophy. Further in vivo studies are needed to better engraftment efficiency and unravel the effects of host environment of transplanted cells. This could be accomplished by varying the age/ disease model of the host. Although several studies of in vitro skeletal muscle microtissues have been described, mostly using C2C12 mouse myoblast cell line, hiPSC-derived 3D microtissues are rare. A study published earlier this year by Rao et al was the first to use hiPSc-derived myocytes in a 3D culture system; however, their method of inducing myogenic differentiation involved ectopic overexpression using a lentivirus. By contrast, our study uses myogenic progenitors that have been derived using small molecule and growth factor cues. Although the introduction of a transgene is not a worry in this case, our method allows for the appearance of Pax7 positive cells, and we show that this satellite cell-like population arises within the microtissue as well. Our method thus sets the stage for in vitro microfluidic studies that can assess satellite cell function and behavior.

Since we are able to derive Pax7 positive cells without ectopic overexpression, it would be interesting to do a head-to-head comparison of Pax7+ cells derived by our method

with Pax7+ myogenic progenitors derived by ectopic overexpression of Pax7. Studying the gene expression profiles and the chromatin structure of both these populations of cells would help us better understand the differences between forced differentiation by ectopic overexpression and a more "natural" derivation. Is the binding of Pax7 similar? Or are there differential binding sites due to an epigenetic memory accrued by one cell type and not the other? What are the downstream functional outcomes of inducing terminal differentiation in these two cell populations? Are both populations equally capable of asymmetric division, since that is a key property of stem cells? Satellite-like cells present an interesting model for this type of study. One concern could be that ectopic overexpression versus directed differentiation yield cells of varying maturity, so comparison of the two would not make sense. However, satellite cells can become quiescent; this has been observed in vitro through ki67 staining (Chal 2015); thus, study of two populations of quiescent satellite-like cells would then control for maturity.