UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Transcriptional and Epigenetic Dynamics Underlying Direct Cardiac Reprogramming

Permalink
https://escholarship.org/uc/item/2ht942dr

Author
Stone, Nicole

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Transcriptional and Epigenetic Dynamics Underlying Direct Cardiac Reprogramming

by

Nicole R. Stone

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Copyright 2017
by
Nicole R. Stone

Dedico esta disertación a mi familia.

PERSONAL ACKNOWLEDGMENTS

Dr. Deepak Srivastava, thank you for the many opportunities you have given me as a student in your lab. Your enthusiasm and support have not wavered throughout this arduous project. Beyond scientific guidance, you have taught me invaluable lessons about how to be a good mentor that I will carry with me for the rest of my life. Deepak created a lab environment filled with brilliant, dedicated, and caring people whom it has been an honor to work with for the past six years, and I thank him for allowing me to be a part of this remarkable group.

I dedicate this work to my family. Mom and Dad, you both worked so hard to allow me the opportunity to pursue my dreams. You gave me everything I need to succeed, but most importantly you gave me the assumption that nothing is out of reach. Thank you to my sister Brenda, for unfailingly being a model of independence, strength, and compassion. Thank you to my brother Brian, who always pushed me to, “Just try it, Nut. Figure it out!” You opened my eyes to music, art, Chinese food, and you helped me understand there was more to the world than our small town. You are all the foundation of who I am today.

Prior to graduate school, many wonderful scientists took on the role of mentor and guided me down the path I took to UCSF. Wade Hazel, thank you for welcoming me into your lab as a high school student and showing me that research should be a fun job. Without you, I would probably have a retirement fund by now, but let’s face it – I

was meant to be a scientist. Your support has opened many doors, many of them within myself. Larry Brody and Faith Pangilinan, thank you for pushing me to set my sights on the top tier. Thank you, Praveen Sethupathy, for taking me under your PERL-y wing. Your gentle introduction to coding empowered me to take on the challenging project I am now completing, and your kindness to everyone in your path inspires me to try for the same.

Every PhD graduate I know describes their degree process as the most challenging experience of their life, and I am no exception. The list of people that helped me persevere includes both scientific mentors and personal lifelines. Katie Pollard, in addition to your guidance as a committee member, your lab has been a haven for me as I transitioned away from the bench and into the analysis phase of my graduate work. I've learned so much from you and from the talented people you've chosen to fill your lab. Kathy Ivey, you were my scientific savior during the most difficult years of this degree. Thank you for always making the time to help in some way, whether that was by offering some technical nugget of wisdom or a stiff drink and a kind ear. Casey Gifford, as a scientist, through your example and your unrelenting pressure, you have pushed this project and myself to both grow into something better. As a loyal friend, you have stood by me through some difficult times, and I thank you for both. Bethany Taylor, thank you for always having an extra laser pointer or pep talk on hand. Karishma Pratt, thank you for all your hard work on this project and for teaching me the joy of mentorship. Seeing you grow as a scientist has been one of the most rewarding things I have experienced.

Finally, I would like to acknowledge some of the friends who became my San Francisco family, in chronological order of our acquaintance. Marie La Russa, your empathy knows no bounds. I am so grateful that our friendship has flourished and endured since our first interview at UCSF – we made it! Ben Cohn, you helped keep music and perspective in my life. Samantha Hastie, when I needed you most, you rescued baby sea turtles with me, and I am forever grateful. Christina Theodoris, thank you for being my lab-sister, partner in late-night science, and sharer of Mom care packages. Sarah Elmes, you made cell sorting something to look forward to, and taught me about what it means to be strong. Tara Friedrich, our friendship has made the worst times better, and the best times also better. Alex Williams I cannot thank you enough for the time you've spent teaching me to debug my code and nurture my caffeine addiction. Nathan Lozier, you are simply the kindest person I know.

Thank you all.

CONTRIBUTIONS AND PUBLICATIONS

The work presented here is the result of years of effort by many dedicated scientists. My primary collaborators have been Casey Gifford and Reuben Thomas. Casey Gifford sequenced and analyzed all single-cell RNA-seq data, and she provided mentorship for the project as a whole. Reuben Thomas completed or taught me how to complete much of the analysis shown here. Reuben also applied a machine learning approach to model the interaction between transcription factors and transcriptional changes during reprogramming, which will be included for publication but is not shown here. I have had the pleasure of working with many other talented scientists over the course of this project, including Tara Friedrich, Sean Thomas, and Alex Williams in the Gladstone Bioinformatics Core. Kaitlen Samse generated ChIP-seq data used to make Figure 4. Karishma Pratt assisted with mouse breeding, cell culture, and generated the ATAC-seq data used to make Figure 3. Amelia Schricker assisted with mouse breeding and cell culture.

Deepak Srivastava, Kathryn Ivey, Casey Gifford, and Katherine Pollard provided conceptual and technical guidance invaluable to the success of this project and will be authors on the forthcoming manuscript. Katherine Pollard was also immensely helpful as a committee member, along with Joseph Costello and Benoit Bruneau, all of whom provided sage advice and constructive criticism.

The data presented here is unpublished, but is currently being prepared for submission:

Stone, N.R., C. A. Gifford, R. Thomas, T. M. A. Mohamed, P. Yu, K. B. Pratt, E. M. Radzinsky, K. Samse, K. N. Ivey, K. S. Pollard, and D. Srivastava. “Transcriptional and Epigenetic Dynamics Underlying Direct Cardiac Reprogramming.” (Manuscript in preparation)

Although not included in this dissertation, I have also contributed data and expertise to the following projects:

(1) Mohamed, T. M., **N. R. Stone**, E. C. Berry, E. Radzinsky, Y. Huang, K. Pratt, Y. S. Ang, P. Yu, H. Wang, S. Tang, S. Magnitsky, S. Ding, K. N. Ivey and D. Srivastava (2017). "Chemical Enhancement of In Vitro and In Vivo Direct Cardiac Reprogramming." Circulation 135(10): 978-995.

(2) Ang, Y. S., R. N. Rivas, A. J. Ribeiro, R. Srivas, J. Rivera, **N. R. Stone**, K. Pratt, T. M. Mohamed, J. D. Fu, C. I. Spencer, N. D. Tippens, M. Li, A. Narasimha, E. Radzinsky, A. J. Moon-Grady, H. Yu, B. L. Pruitt, M. P. Snyder and D. Srivastava (2016). "Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis." Cell 167(7): 1734-1749.e1722.

(3) Fu, J. D., **N. R. Stone**, L. Liu, C. I. Spencer, L. Qian, Y. Hayashi, P. Delgado-Olguin, S. Ding, B. G. Bruneau and D. Srivastava (2013). "Direct Reprogramming of Human Fibroblasts toward a Cardiomyocyte-like State." Stem Cell Reports 1(3): 235-247.

(4) Yu, P., C. Stilson, A. Chin, E. Radzinsky, Y. Huang, A. Schricker, R. Krug, S. Metzler, **N. R. Stone**, K. Pratt, A. Liang, T. Moore, T.M. Mohamed, I. Liu, M. Saeed, and D. Srivastava. "In vivo direct cardiac reprogramming with human transcription factors in a pig model." (Manuscript in preparation)

(5) *Yu, P., *Ojala, D., Y. Huang, E. Radzinsky, A. Schricker, **N. R. Stone**, C. A. Gifford, K. Pratt, D. Schaffer, and D. Srivastava. "A novel AAV variant with fibroblast tropism successfully facilitates direct cardiac reprogramming." *These authors contributed equally. (Manuscript in preparation)

Transcriptional and Epigenetic Dynamics Underlying Direct Cardiac Reprogramming

By

Nicole R. Stone

Abstract

Despite numerous advances, direct reprogramming of fibroblasts to induced cardiomyocytes remains an inefficient process with largely unknown underlying mechanisms. To understand the molecular events associated with direct cardiac reprogramming, we characterized genome-wide transcriptional changes in individual and bulk reprogramming cell populations, examined genomic changes in chromatin accessibility and DNA methylation levels during reprogramming, and mapped DNA binding by Gata4, Mef2c, and Tbx5 (GMT) that initiates the transition from fibroblast to induced cardiomyocyte. We discovered that the transition to cardiac-like gene expression occurs within 48 hours of initiating transdifferentiation. This change is accompanied by rapid epigenomic remodeling and includes dynamic chromatin regions enriched for transcription factor motifs in addition to GMT, with different motif families associated with specific patterns of chromatin and DNA methylation changes. We found evidence for binding of the reprogramming factors to dynamically-opening distal regulatory regions linked to cardiac function. In addition, we document the mechanism through which TGFβ and Wnt signaling pathway inhibition improves reprogramming efficiency by limiting aberrant reprogramming outcomes. These results provide insight into the mechanisms that drive direct cardiac reprogramming and provide a hypothesis-building framework for improving this process.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION……1

1.1 Reprogramming cell fate……2

1.2 Epigenetic remodeling during cell fate transitions……3

1.3 Direct reprogramming……4

1.4 Direct cardiac reprogramming……5

1.5 What we are contributing……6

CHAPTER 2: RESULTS……8

2.1 Heterogeneous response to reprogramming……9

2.2 Mechanisms of reprogramming improvement by TGFβ and Wnt signaling inhibition……16

2.3 Chromatin accessibility dynamics reveal transcription factor motifs enriched in different aspects of reprogramming……22

2.4 Chromatin occupancy by Gata4, Mef2c, and Tbx5 reveals regions important for reprogramming……26

2.5 DNA methylation dynamics associated with reprogramming……34

CHAPTER 3: DISCUSSION……37

3.1 Single-cell RNA-seq reveals mechanisms of reprogramming……38

3.2 Chromatin accessibility and DNA methylation dynamics highlight reprogramming complexity……39

3.3 Conclusions……42

3.4 Future directions……42

CHAPTER 4: METHODS……..45

REFERENCES……..53

LIST OF FIGURES

Figure 1. Single-cell expression analysis of direct cardiac reprogramming.....……..12

Figure 1.1. Cluster components of single-cell expression analysis……...……..……14

Figure 1.2. Bulk RNA-seq shows distinct patterns of transcriptional change.....……15

Figure 2. TGFβ and Wnt inhibition promotes cardiac identity and suppresses aberrant reporter activity…….……...……..……19

Figure 2.1. TGFβ and Wnt inhibition alters early reprogramming trajectory..……...….21

Figure 3. Direct cardiac reprogramming initiates widespread chromatin accessibility changes…….....……..……25

Figure 4. Chromatin accessibility dynamics associated with Gata4, Mef2c, and Tbx5 occupancy…….……...……..……29

Figure 4.1. Biological processes associated with Gata4, Mef2c, and Tbx5……...30

Figure 4.2. Regions bound by Gata4, Mef2c, and Tbx5, individually and in combination, display different chromatin accessibility trends……...……..……32

Figure 5. DNA methylation changes associated with reprogramming..……..……..……36

CHAPTER 1
INTRODUCTION

1.1 Reprogramming cell fate

Cellular differentiation, a basic component of embryonic development and mature tissue maintenance, requires precise regulation of gene expression and underlying chromatin structure. Historically, cellular differentiation was considered unidirectional, with immature cells following an irreversible path towards progressive lineage commitment. However, this concept was challenged by the discovery of somatic cell nuclear transfer, the precursor of cellular reprogramming as we know it today. This process involved harvesting a nucleus from a fully differentiated cell in a tadpole's intestinal tract and inserting it into an enucleated oocyte, resulting in a new cell that was capable of forming all cell types in a new tadpole's body (Gurdon 1962).

Another conceptual leap occurred with the discovery that expression of a single transcription factor, rather than replacement of an entire nucleus, is sufficient to facilitate cell fate conversion. Ectopic expression of the myogenic transcription factor *MyoD* is sufficient to convert fibroblasts to myoblasts, and to induce various somatic cell types to express myogenic genes, in a process called "reprogramming" (Davis, Weintraub et al. 1987, Weintraub, Tapscott et al. 1989). However, for most cell types no single factor was sufficient to induce cellular reprogramming to a specific lineage. The recognition that a combinatorial transcription factor code may be required for reprogramming paved the way for reprogramming somatic cells towards various alternatives fates, including the pluripotent state, which is achieved by the introduction of Oct4, Sox2 Klf4 and c-Myc (Takahashi and Yamanaka 2006). The ability of induced pluripotent stem cells (iPSCs) to differentiate into any cell lineage further demonstrated the amazing malleability of cellular identity.

1.2 Epigenetic remodeling during cell fate transitions

Cellular reprogramming intrinsically involves epigenetic remodeling. To induce a cell fate change, widespread repatterning of chromatin structure from that of the starting cell type to the target cell type must occur. Repeatedly, we find a dynamic epigenetic landscape underlies cell fate changes such as development, reprogramming, and tumorigenesis (Wamstad, Alexander et al. 2012, Gifford, Ziller et al. 2013, Papp and Plath 2013, Wapinski, Vierbuchen et al. 2013, Aldiri, Xu et al. 2017, Wapinski, Lee et al. 2017). Chemical modifications to histones, the proteins which organize and package DNA in the nucleus, is one category of epigenetic regulation fundamentally linked to chromatin state and cell identity. Different combinations of histone modifications mark promoters and enhancers, and are correlated with transcriptional activation, repression, and a “poised” state associated with genes transitioning to an active state (Guenther, Levine et al. 2007, Heintzman, Stuart et al. 2007, Creyghton, Cheng et al. 2010, Ernst, Kheradpour et al. 2011). During differentiation from iPSC to cardiomyocyte, there is a coordinated relationship between histone modification-based chromatin signatures and expression of functionally-related genes, and these patterns identify distinct stages of differentiation and epigenetic priming of repressed cardiac loci (Wamstad, Alexander et al. 2012). More recently, profiling of chromosomal architecture during cardiomyocyte differentiation revealed coordinated three-dimensional reorganization, including decompaction of chromatin regions containing cardiac-specific genes which were transcriptionally activated during differentiation (Fields 2017).

In addition to transcriptional activation of genes in the target cell type, silencing of genes expressed in the starting population is crucial for stable transition to a new cell

state. Relative to histone modifications, DNA methylation is considered a more permanent epigenetic change, and is generally associated with transcriptional silencing. During iPSC reprogramming, changes in DNA methylation happen only after changes in other epigenetic markers have already occurred (Polo, Anderssen et al. 2012). Similarly, this order of events is observed during differentiation from iPSCs to various lineages, where silencing by histone modifications occurs prior to deposition of DNA methylation (Gifford, Ziller et al. 2013). Within normal postnatal cardiac development, dynamic DNA methylation is associated with silencing of the fetal gene expression program, facilitating myocyte maturation to fully functional adult cardiomyocytes (Gilsbach, Preissl et al. 2014).

1.3 Direct reprogramming

While the differentiation of induced pluripotent stem cells created a robust platform for disease modeling and a potential source of transplantable tissues, there are many reasons why this process is not ideal for therapeutic purposes. Some of these barriers include: (1) iPSC generation is an inefficient and slow process, (2) differentiation to mature cells types is rarely robust, (3) incomplete or inappropriate differentiation can lead to teratoma formation, and (4) survival and integration of transplanted cells is limited (Hong, Winkler et al. 2014, Kimbrel and Lanza 2015, Garcia-Castro and Singec 2017). The necessity of seamless integration is especially crucial for cardiomyocytes, which must electrically couple to neighboring cells for proper function *in vivo*. Human embryonic stem cell-derived cardiomyocytes transplanted into primate hearts improve function, but lead to arrhythmia (Chong, Yang et al. 2014).

To circumvent the need for reversion to pluripotency, combinations of transcription factors that could facilitate direct conversion between somatic cell types were sought as an alternative. To that end, essential regulators of development were employed to create populations that mirror the desired cell type on both a molecular and phenotypic level. Using this strategy, direct conversion of somatic cells to a wide variety of cell types was accomplished, including macrophages, adipocytes, hepatocytes, pancreatic β-cells, neuronal subtypes, and cardiomyocytes (Feng, Desbordes et al. 2008, Kajimura, Seale et al. 2009, Ieda, Fu et al. 2010, Vierbuchen, Ostermeier et al. 2010, Huang, He et al. 2011, Sekiya and Suzuki 2011, Vierbuchen and Wernig 2012).

1.4 Direct cardiac reprogramming

In 2010, our lab reported successful *in vitro* direct conversion of mouse fibroblasts to induced cardiomyocyte-like cells (iCMs), without transition through a pluripotent stage, using ectopic expression of three transcription factors crucial to cardiac development: Gata4, Mef2c, and Tbx5 (Ieda, Fu et al. 2010). iCMs created by direct cardiac reprogramming are defined by activation of cardiomyocyte genes, repression of fibroblast genes, formation of sarcomere structures, calcium transients, and, rarely, spontaneous contraction. This transdifferentiation process occurs directly, without transitioning through a cardiac precursor state. Building upon this, *in vivo* studies showed that direct cardiac reprogramming can result in reduced scar size and functional improvements in cardiac output in a mouse model of infarction (Qian, Huang et al. 2012). Supplementing with two additional transcription factors, Mesp1 and Esrrg, supports direct reprogramming in human fibroblasts, demonstrating the potential of this approach for therapeutic usage (Fu, Stone et al. 2013).

Screening of transcription factors, non-coding RNAs, and drugs have led to improvements in direct cardiac reprogramming outcomes. These variations include: modulating the stoichiometry of reprogramming factors, expressing additional factors including Hand2, Myocd, *miR-1, and miR-133*, and modulation of signaling pathways including Akt/protein kinase B, TGFβ, Wnt, and Notch (Nam, Song et al. 2013, Wang, Liu et al. 2015, Wang, Liu et al. 2015, Zhou, Dickson et al. 2015, Abad, Hashimoto et al. 2017, Mohamed, Stone et al. 2017). Along with improving reprogramming efficiency, these breakthroughs provide some insight into the signaling pathways and molecular mechanisms that are crucial for reprogramming success. The findings that removal of epigenetic repressor Bmi1 and chemical depletion of suppressive histone modifications leads to improved reprogramming outcomes highlights the importance of tightly regulated epigenetic remodeling for successful reprogramming to occur (Hirai and Kikyo 2014, Zhou, Wang et al. 2016). Indeed, profiling of histone modifications H3K27me3 and H3K4me3 during direct cardiac reprogramming revealed an early switch from a repressed to an active epigenetic state at cardiac promoters, followed by progressive repression at fibroblast promoters occurring one week later (Liu, 2016). However, despite these advances, iCM maturation outcomes remain inconsistent, efficiency remains low, and basic mechanisms behind direct cardiac reprogramming remain unknown.

1.5 What we are contributing

We hypothesized that comprehensive evaluation of the shifting transcriptional and epigenetic landscapes during cardiac reprogramming would uncover essential regulators of this conversion process, define the stages that a reprogramming cell

passes through during conversion to an iCM, and reveal the epigenetic contexts through which reprogramming factors function. Here, we provide a comprehensive view of the transcriptional and epigenetic dynamics that occur during cardiac reprogramming using the combination of Gata4, Mef2c and Tbx5 *in vitro*. Single-cell RNA sequencing reveals that a cardiac fate is acquired within 48 hours while maturation continues for 2 weeks in the absence of proliferation. Profiling of open chromatin and DNA methylation levels confirm that the majority of epigenomic remodeling events occur early and reveals important roles for TEAD, H-box, MAD, bHLH, and ETS factors.

From this work, we conclude that Gata4, Mef2c, and Tbx5 initiate a rapid transcriptional and epigenetic transformation in cardiac fibroblasts. Dynamic chromatin regions were enriched for transcription factor motifs in addition to GMT, with different motif families associated with specific patterns of chromatin and DNA methylation changes. We report evidence for binding of the reprogramming factors to dynamically-opening distal regulatory regions linked to cardiac function. In addition, we document the mechanism through which TGFβ and Wnt signaling pathway inhibition improves reprogramming efficiency by promoting cardiac identity and limiting aberrant reprogramming outcomes, providing insight into the mechanisms that drive direct cardiac reprogramming.

CHAPTER 2
RESULTS

2.1 Heterogeneous response to reprogramming

We induced direct cardiac reprogramming of $Thy1^+$ cardiac fibroblasts isolated from neonatal heart tissue from the αMHC-GFP reporter mouse line. Fibroblasts were directly reprogrammed to GFP^+ induced cardiomyocytes (iCMs) using one of two reprogramming protocols: infection with retroviruses to force expression of (1) Gata4, Mef2c, and Tbx5 (GMT) alone or (2) with addition of TGFβ and Wnt chemical inhibitors (GMTc) (Figure 1A). To identify transcriptional states present during direct cardiac reprogramming, we generated 25,970 single-cell transcriptomes from cells collected at days 1, 3, 7, and 14 days following initiation of GMT or GMTc reprogramming protocols (Figures 1B). We identified 18 transcriptionally distinct clusters that correlate to several different cell types and states (Figures 1C, 1D). Importantly, technical variability was limited as all clusters are comprised of cells isolated at multiple time points (Figure 1.1A) and each time point segregated into multiple clusters (Figure 1.1B).

Marker genes within each cluster give insights into cluster identities. Expression of connective tissue growth factor (*Ctgf*) marks cells most similar to fibroblasts, including the starting population of fibroblasts (Day -1; clusters 3 and 13) and cells which have attained an intermediate or alternative reprogramming state, primarily clusters 2, 5, 10, 11 (Figure 1D-E). Cardiac troponin T (*Tnnt2*) and myosin light chain 7 (*Myl7*), two genes highly expressed in cardiomyocytes, mark the “most reprogrammed” iCM populations clusters 1, 8, and 11 (Figure 1D-E). Expression of capthesin S (*Ctss*), a lysozymal cysteine protease, marks blood- or immune cell-related clusters 6, 7, 9, and 12 (Figure 1D-E). Clusters 15 and 16 contain cells expressing markers of endocardium and motor neurons, respectively, likely contaminants from fibroblast isolation as they are present in

the starting population (Figure 1D). We also observed that cluster 4 contains all reprogramming cells that remain proliferative, which supports previous data suggesting that cell cycle exit is required for cardiac reprogramming (Figure 1.1C).

In addition, our analysis reveals divergence from fibroblasts (Day -1) into one of three possible states at the earliest time point of cells in which reprogramming has been initiated (Day +1) (Figure 1B). Separation of cells subjected to different reprogramming protocols is apparent at Day +3, the earliest time point at which we compare protocol outcomes, and continues at Day 7 (Figure 2.1A-B). The absence of small molecules produces cluster 5, while GMTc causes association with cluster 0, with both alternative reprogramming outcomes expressing both fibroblast and cardiomyocyte genes (Figure 1C-E). While each reprogramming protocol produces a unique alternative trajectory according to our analysis, a similar reprogramming trajectory is observed with cells in "most reprogrammed" clusters 1 and 8 being derived from both protocols. This suggests that the rate at which cells acquire a cardiomyocyte-like transcriptional signature is similar, but they differ in the percent of cells that enter the state; 27.9% of GMTc cells collected at day 7 are associated with cluster 8, compared with only 3.8% of GMT cells (Figure 1B, S1A-B).

Within the most CM-like clusters (1, 8, and 11), separation is driven in part by the cells' ability to repress the fibroblast gene expression program. Genes that distinguish cluster 8 from clusters 1 and 11 include insulin-like growth factor-binding protein 7 (*Igfbp7*), *Nupr1,* and *Col5a2* which are not associated with the iCM-like transcriptional state (Figure 1F-G).

In summary, we show that direct cardiac reprogramming is a highly heterogeneous process. Surprisingly, we found that cells expressing our αMHC-GFP reporter of cardiac reprogramming consisted of not only iCMs, but also non-cardiac and intermediately reprogrammed cell types. Heterogeneity in fibroblast response is apparent as early as day 1 of reprograming induction. Additionally, the rate at which GMT- and GMTc-induced iCMs occur is similar, but outcomes from the two protocols differ in the proportion of fibroblasts that become iCMs rather than taking alternative trajectories. Additionally, we confirm that successfully reprogrammed iCMs rapidly exit cell cycle.

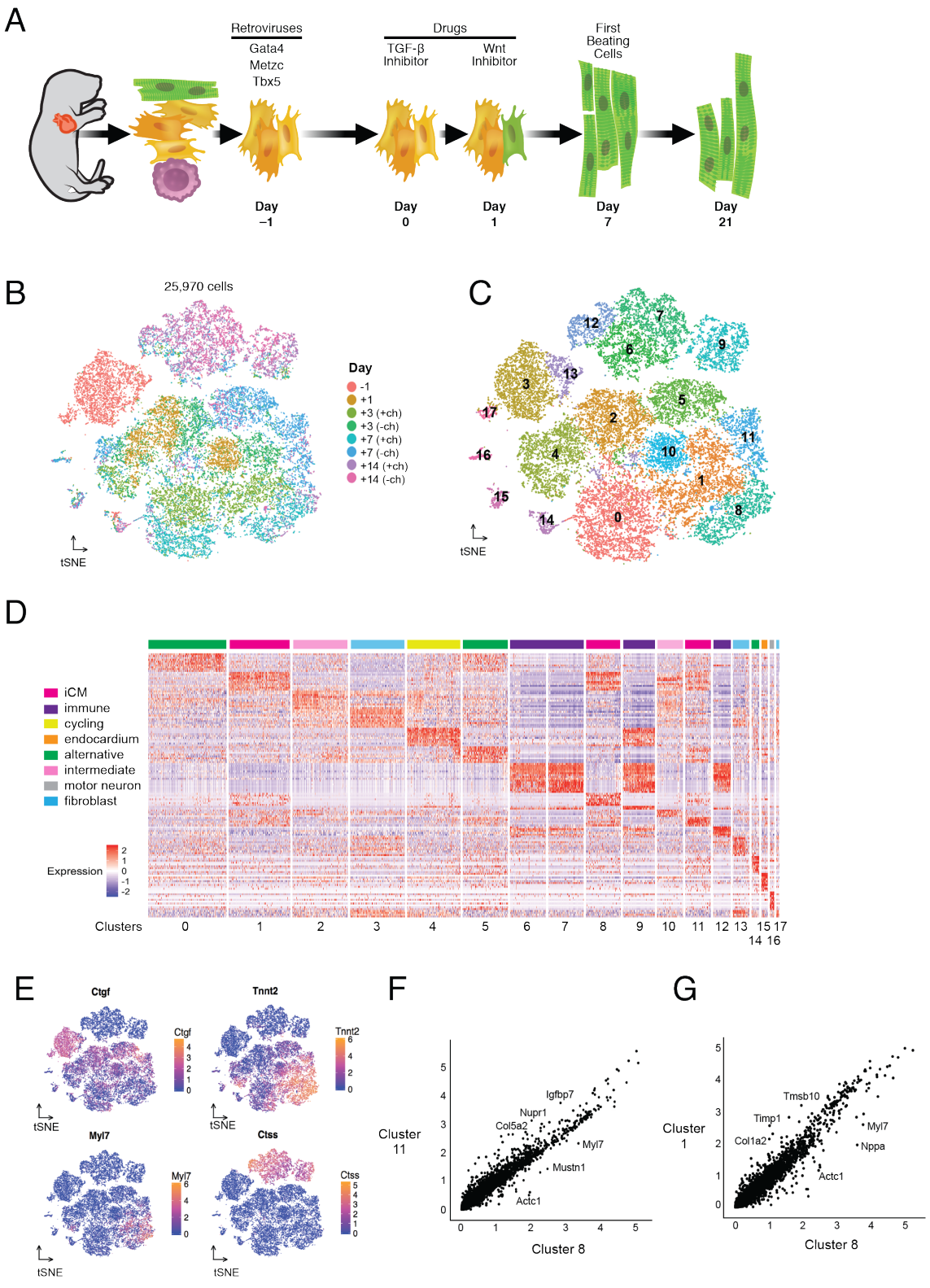


Figure 1. Single-cell expression analysis of direct cardiac reprogramming. (A) Schematic of system. Tissue explants from αMHC-GFP^{+} neonatal mouse pups are

cultured for ~10 days, allowing expansion of fibroblast population. Thy1$^+$ fibroblasts are isolated by FACS, infected with GMT retroviruses, and cultured with or without chemical inhibitors for up to three weeks as some reprogrammed cells mature into spontaneously beating iCMs. (B) tSNE plot of 25,970 cells collected at five time points from two different reprogramming conditions. Each cell is one dot and the distance relationship between the cells was calculated using statistically significant principal components. The color indicates the day of collection and condition. Within the legend, “+ch” indicates chemical inhibitors were added, and “-ch” indicates no chemical inhibitors were added. (C) tSNE plot of 18 clusters identified by principal component analysis and a graph-based clustering approach. Clusters are numbered and distinguished by color. (D) Heatmap including top 15 marker genes identified for each cluster. Each row is 1 gene and each column is 1 cell. Red indicates high expression; blue is low. A legend of cell identities for each cluster is at left, and indicated above the cluster. (E) Expression level of key marker genes. Orange indicates high expression; blue is low. (F) Scatterplot of expression levels of genes identified as markers of clusters 8 and 11. Differentially expressed genes are indicated. (G) Scatterplot of expression levels of genes identified as markers of clusters 8 and 1.

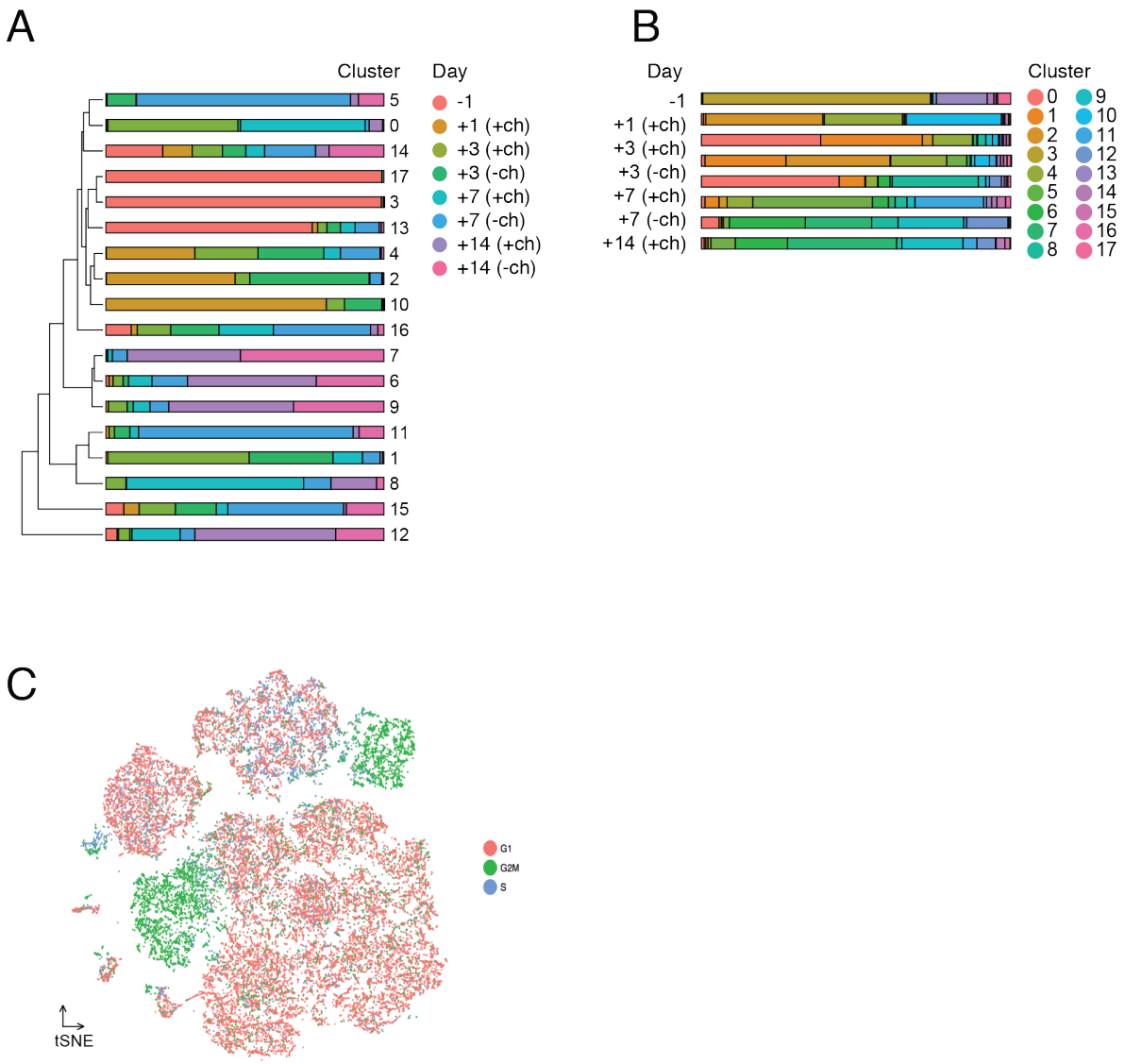


Figure 1.1. Cluster components of single-cell expression analysis. (A) Boxplot where each box illustrates the sample contribution to each cluster. Dendrogram was calculated using principal component analysis and conveys relationship between clusters. Chemical inhibitor treatment indicated by “+ch”; lack of chemical treatment indicated by “-ch”. (B) Boxplot where each box illustrates one sample and the color indicates the cluster. Legend at right. (C) tSNE plot of cells colored by cell cycle identity.

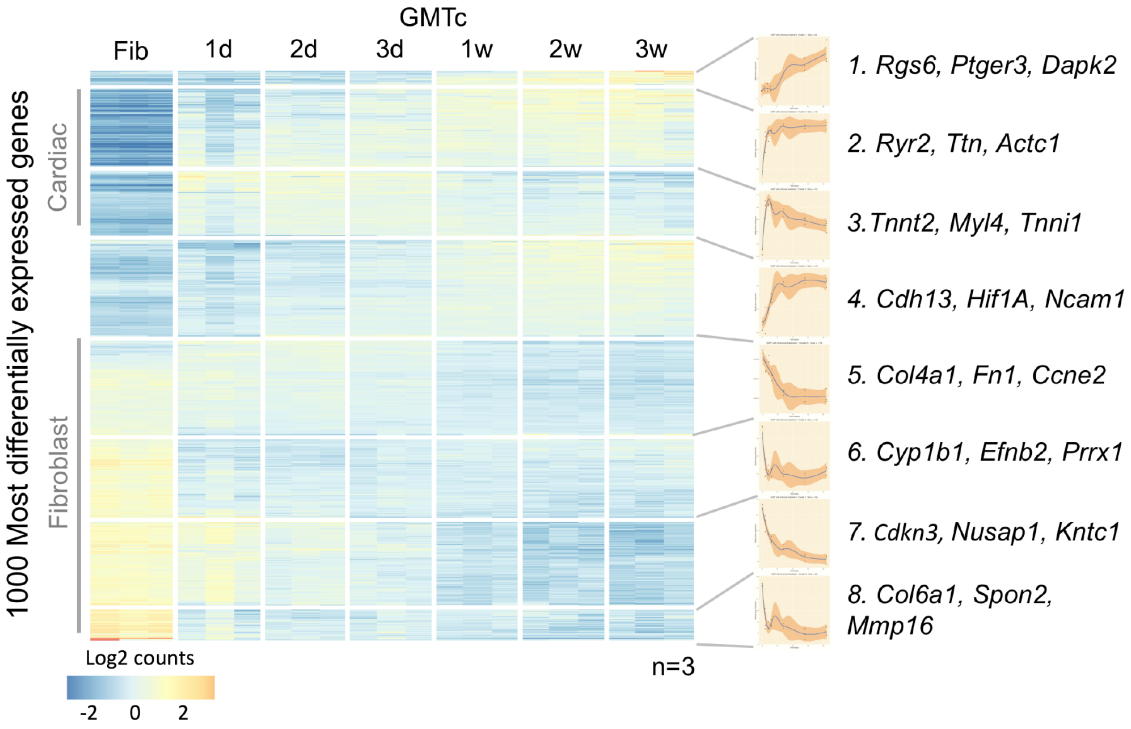


Figure 1.2. Bulk RNA-seq shows distinct patterns of transcriptional change. Heatmap showing expression levels from bulk RNA-seq of 1000 most differentially expressed transcripts between fibroblasts and GFP^{+} sorted iCMs at timepoints during GMTc reprogramming from day 1 (1d) up to 3 weeks (3w). Color scale shown at bottom. Each time point was collected in triplicate, and all replicates are shown here. Medoid plots of representative expression trends shown in orange. Example genes from each cluster listed at right.

2.2 Mechanisms of reprogramming improvement by TGFβ and Wnt signaling inhibition

Since previous evidence suggests that TGFβ and Wnt signaling inhibition results in a more functional iCM, we next wanted to understand how this signaling inhibition facilitates reprogramming. We collected day 14 cells treated with GMT or GMTc, sorted αMHC-GFP$^+$ iCMs, then compared the single cell transcriptomes of 12,905 sorted and unsorted cells (Figure 2A-B). We found that reporter-based sorting isolates a cell population generally expressing cardiac marker genes including cardiac troponin T (*Tnnt2*) without expressing blood cell marker genes including *Ctss* (Figure 2B-D), regardless of protocol. Meanwhile, most unsorted cells subjected to either protocol generally express *Ctss* without expression of *Tnnt2* (Figure 2A, 2c-d).

Among these cells, we identified 7 transcriptionally distinct clusters using PCA (Figure 2E-F). Cluster 1 represents our “most reprogrammed” cluster, and is defined by expression of cardiac genes including *Actc1* and *Myl7*. Cells in cluster 4 express high levels of fibroblast genes including *Fn1* and *Col1a2*, suggesting that this cell population is merely reporter positive without actually undergoing reprogramming. Cluster 6 exhibits high expression of desmin (*Des*) an intermediate-filament protein expressed in both cardiac and skeletal muscle, and mesothelin (*Msln*), an EMT regulator, suggesting this group of cells has responded to GMT expression by following an alternative reprogramming trajectory, unlike the majority of iCMs which fall into cluster 1.

Clusters 0, 2, and 3 are defined by expression of non-cardiac cell markers (Figure 2E-F). Further delineating cell groups within these non-cardiac trajectories, Cluster 0 expresses two markers of the blood cell lineage, *Spp1* and *Lyz2* (S2C).

Cluster 2 expresses marker of proliferation ki-67 (*Mki67*) and *Top2a*, a topoisomerase that repairs double-strand DNA breaks, indicating that cluster 2 contains cells which are actively dividing or have sustained DNA damage. Cluster 3 contains cells expressing matrix metalloproteinase-12 (*Mmp12*), a macrophage gene; and *Fabp4*, a fatty acid binding protein found in adipocytes.

To quantitatively compare the content of each cluster, we plotted the proportion of each sample within each cluster (Figure 2G). Cluster 1, contains the highest proportion of reporter$^+$ cells ("pos"); most cells in this cluster were made using the GMTc protocol ("pos (+ch)"). Clusters 0, 2, 3, and 4 contain few iCMs, and are mainly comprised of unsorted cells from either protocol. Cluster 5 contains similar relative amounts from each sample type. Cluster 6 is made almost entirely of reporter$^+$ cells, with a slightly higher contribution from iCMs derived from the chemical-free protocol (Figure 2G).

Conversely, we also examined the distribution of clustering within each sample type. Both unsorted samples have similar distributions among clusters. However, when we examine clustering of sorted reporter$^+$ cells ("pos"), major differences exist between protocols. The majority of iCMs generated with the GMTc protocol ("pos +") fall into the "most reprogrammed" cluster 1, while iCMs generated without TGFβ nor Wnt inhibition ("pos -") are distributed among all clusters, demonstrating that GFP$^+$ cells generated with the GMTc protocol are more uniformly reprogrammed than those generated with GMT alone (Figure 2H). Alternatively, this difference suggests that aberrant reporter activation in non-reprogramming cells is suppressed by TGFβ and Wnt inhibition.

For comparison, we looked at the 40 most differentially expressed marker genes identified for day 14 non-cardiac clusters 0, 2, and 3 within bulk RNA-seq data from sorted GFP^+ iCMS. Consistent with the single-cell expression data, in the bulk data, these genes are generally most highly expressed at day 7 and day 14 in iCMs generated with the chemical-free protocol. However, iCMs generated with GMTc generally retain expression levels nearer to or lower than in fibroblasts (Figure 2I).

Within iCM cluster 1, chemical inhibitors ("+ch") produce a more uniform expression range for cardiac markers *Tnnt2* and *Myl4*, and suppress expression of Wnt induced signaling pathway protein 1 (*Wisp1*). Within GFP^+ sorted iCM ("pos") populations, cardiac *Tnnt2* and *Myl4* expression distributions are similar regardless of protocol. However, TGFβ and Wnt inhibition decreases expression of *Wisp1* and results in more uniform expression throughout the GFP^+ population (Figure 2J).

In conclusion, we demonstrate that blood marker-expressing cell populations are more prominent at day 14 than earlier time points, possibly indicating delayed expansion of a contaminating cell type not fully eliminated by Thy1 sorting. Any cell population that does not exit the cell cycle would overtake the iCM population over time, since cell cycle exit is necessary for successful reprogramming to occur. In addition, we document aberrant reporter expression in these non-cardiac cells, which is mitigated by TGFβ and Wnt inhibition. Further, reporter-positive cells are more uniformly cardiomyocyte-like when generated with the GMTc protocol than with GMT alone. Altogether, these results point to the importance of multi-marker validation of reprogrammed cell fate, and also demonstrate how GMTc reprogramming is superior to GMT alone.

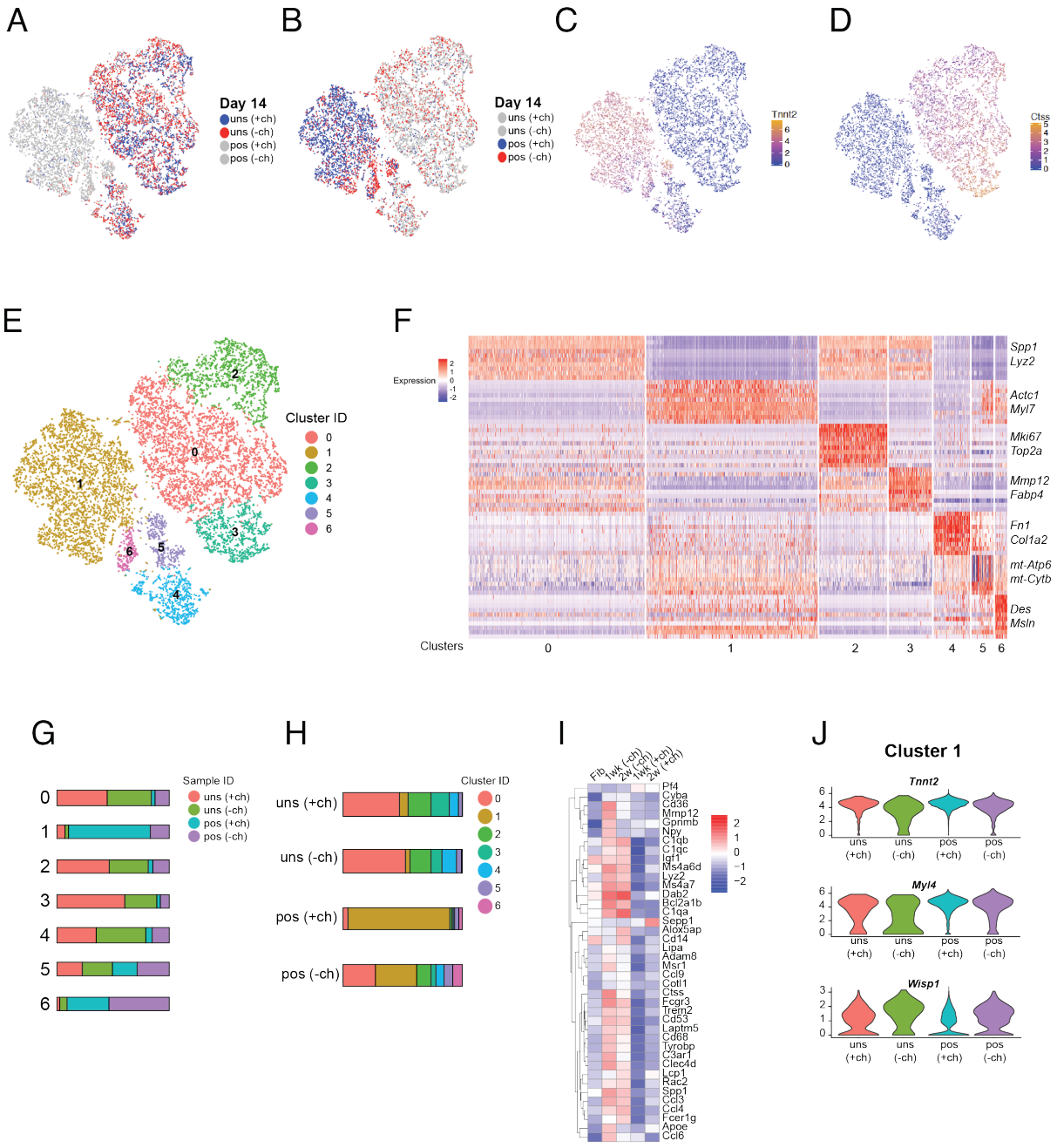


Figure 2. TGFβ and Wnt inhibition promotes cardiac identity and suppresses aberrant reporter activity. (A) tSNE plot including 12,905 cells collected at day 14 of cardiac reprogramming. Uns, unsorted; pos, GFP$^+$ sorted. FACS sorting for GFP$^+$ iCMs was done for 2 samples (pos). Colors highlight cells from unsorted samples. Chemical inhibitor treatment indicated by "+ch"; lack of chemical treatment indicated by "-ch". (B) Same as A, except the colors highlight cells from the GFP$^+$ sorted samples. (C) Expression level of *Tnnt2*, a maker of cardiac identity. Orange indicates high expression while blue is low. (D) Expression level of *Ctss*, a marker of macrophage identity. Orange indicates high expression while blue is low. (E) tSNE plot of 7 clusters identified by

principal component analysis and a graph-based clustering approach. Clusters are numbered and distinguished by color. (F) Heatmap including top 10 marker genes identified for each cluster. Each row is 1 gene and each column is 1 cell. Red indicates high expression while blue is low. Representative marker genes are labeled at right. (G) Boxplot where each box illustrates the sample contribution to each cluster. Color legend is at right. (H) Boxplot where each box illustrates one sample and the color indicates the cluster. Legend at right. (I) Bulk RNA-seq heatmap including row scaled z-scores of gene markers of clusters 0, 2 and 3. (J) Violin plot of expression levels of *Tnnt2, Myl4* and *Wisp1* in cells included in cluster 1.

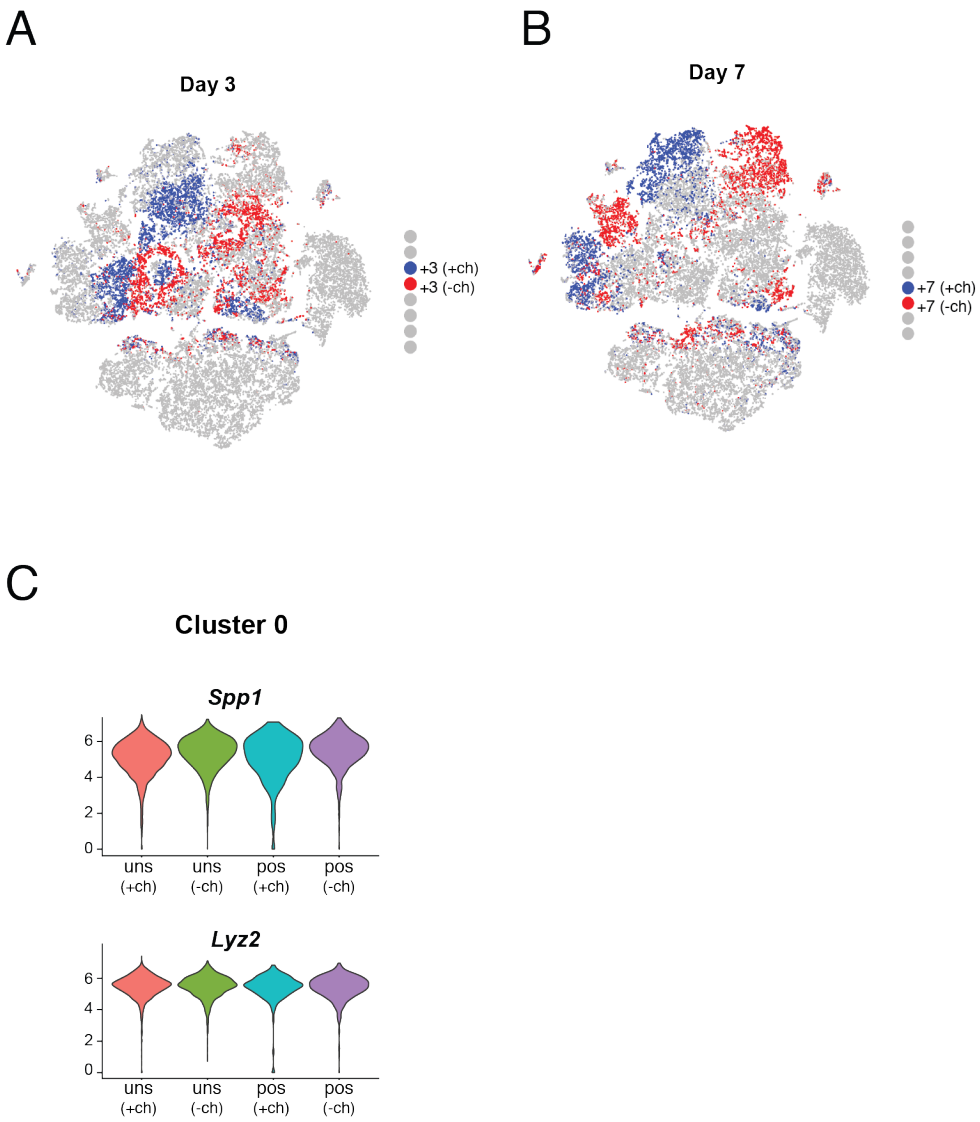


Figure 2.1. TGFβ and Wnt inhibition alters early reprogramming trajectory. (A) tSNE plot including 12,905 cells collected up to day 14 of cardiac reprogramming from two conditions. Unsorted samples collected at day 3 of reprogramming were generated using the GMTc and GMT protocols, indicated by blue and red, respectively. Chemical inhibitor treatment indicated by "+ch"; lack of chemical treatment indicated by "-ch". (B) Same as A, except samples were collected at day 7 of reprogramming. (C) Violin plots of expression levels of *Spp1* and *Lyz2* in cells included in cluster 0.

2.3 Chromatin accessibility dynamics reveal transcription factor motifs enriched in different aspects of reprogramming

To assess chromatin accessibility dynamics, we completed ATAC-seq at time points during GMTc reprogramming from neonatal mouse cardiac fibroblasts to week 3 αMHC-GFP^{+} iCMs. We applied hierarchical clustering to identify multiple patterns of chromatin accessibility variations over our time course, including expected changes such as the gain of accessibility by day 3 of reprogramming in the *Tnnt2* locus which is closed in fibroblasts and remains open at day 7 (Figure 3A). We examined chromatin accessibility dynamics in the 10,000 most differentially open regions identified during our time course, and identified 8 primary patterns of change (Figure 3B).

Within these regions, the most drastic changes occur rapidly, within 48hr of reprogramming induction. More moderate changes occur thereafter, up to week 3. (Figure 3C). Clusters 5 and 6 demonstrate temporary shifts in chromatin accessibility at early time points, with a return towards fibroblast state in the average iCM at later time points (Figure 3C). To highlight transcription factor families that may drive these patterns of change, we analyzed motif enrichment in each cluster. Notably, the T-box family of transcription factor motifs are the most significantly enriched in both region clusters 5 and 6, suggesting the possibility of an early and transient role for reprogramming factor Tbx5 (Figure 3D).

Motif families including Gata4, Mef2c, and Tbx5 motifs, members of Zinc finger (Zf), MADS-box, and T-box families, respectively, are enriched in all 8 clusters of highly dynamic regions, including both opening and closing chromatin regions (Figure 3D). This suggests that these reprogramming factors may be directly involved in both

activation and repression of regulatory loci. Several other transcription factor families have motifs enriched in both opening and closing regions. Basic helix-loop-helix (bHLH) factor motifs are enriched in clusters undergoing both opening (clusters 6-8) and closing (cluster 1). However, the MAD family was enriched in clusters closing, while MADS family motifs are enriched in clusters opening. TEA family motifs were enriched only in clusters exhibiting a closing chromatin trend (clusters 1-4).

To assess the potential functional roles of each cluster, we used nearby gene annotations to assigned biological meaning to these chromatin accessibility changes (Figure 3E). Regions associated with decreasing chromatin accessibility, clusters 1-4, are linked to general and non-cardiac-specific processes such as protein localization, peptide cross-linking, chemokine signaling, and smooth muscle adhesion. Clusters 2 and 4, both associated with closing chromatin, are functionally linked to terms involving immune function, including inflammatory response, leukotriene production, and monocyte chemotaxis. Clusters 5, 7, and 8, associated with opening chromatin during reprogramming, are overwhelmingly linked to cardiac muscle related biological process terms, supporting a chromatin landscape remodeled towards a more cardiomyocyte-like state (Figure 3E).

These trends support the changes in gene expression seen in bulk RNA-seq, which show rapid induction of cardiac-related gene expression and suppression of fibroblast and general proliferation-related genes by day 1 (Figure 1.2). Bulk ATAC-seq dynamics also highlight chromatin-based suppression of immune identity in the GFP^{+} iCM population generated using TGFβ and Wnt inhibition, further supporting the immune cell repressive role of these chemicals demonstrated by single-cell RNA-seq.

In summary, drastic changes in chromatin accessibility occur by day 2 of cardiac reprogramming as measured by ATAC-seq. Reprogramming factor motif enrichment is present within dynamic chromatin regions both gaining and losing accessibility, suggestive of direct GMT influence on both activation and repression of loci. Transcription factor families outside of GMT are also linked to dynamic chromatin regions, including bHLH, H-box, and NR motifs, which are enriched in both dynamically-opening and closing regions, as well as TEA and MAD motifs, which are enriched only in closing chromatin regions. In addition, closing chromatin regions are associated with blood cell-related functional terms, suggesting that TGFβ and Wnt inhibition may promote silencing of these loci by a GMT binding-mediated mechanism. Altogether, the bulk ATAC-seq data reveals motifs and transcription factors that may be responsible for promoting the cardiac state and/or inhibiting other, alternative cell types throughout the reprogramming process.

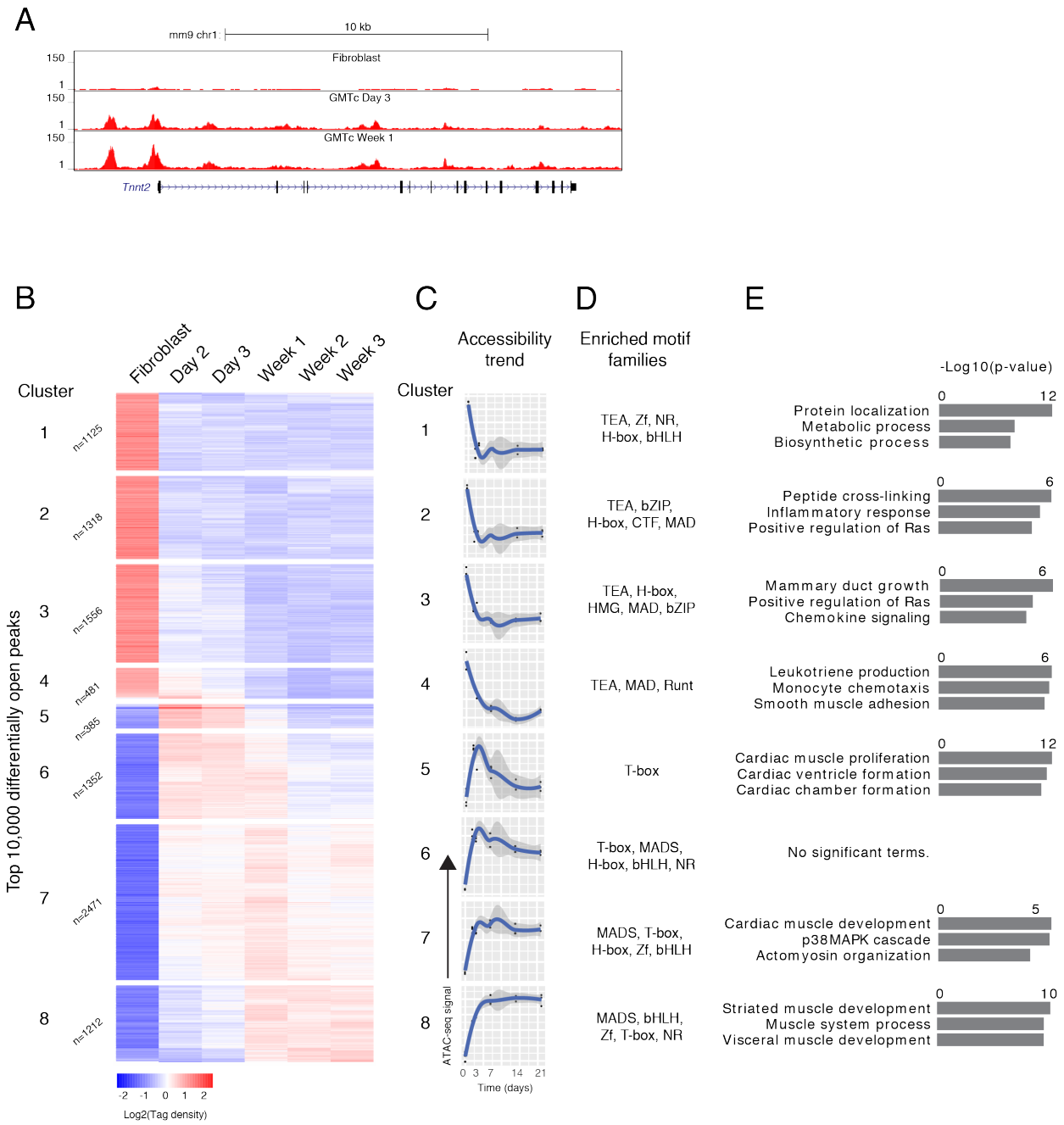


Figure 3. Direct cardiac reprogramming initiates widespread chromatin accessibility changes. (A) Browser tracks display a gain in dynamic ATAC-seq signal in iCMs compared with fibroblasts near cardiomyocyte gene locus, *Tnnt2*. Track height is normalized to sample read depth. (B) Heatmap summarizing HOPACH-clustered chromatin state trends in top 10,000 most dynamic ATAC-seq peaks from GMTc reprogramming time course. (C) Medoid plots of ATAC-seq tag density over time at dynamic regions representative of cluster trends. (D) Transcription factor families with motifs significantly enriched within region clusters. (E) Biological process terms enriched in dynamic ATAC-seq peak clusters.

2.4 Chromatin occupancy by Gata4, Mef2c, and Tbx5 reveals regions important for reprogramming

To understand how the three reprogramming transcription factors drive direct cardiac reprogramming, we performed ChIP-seq on unsorted cells at an early stage of reprogramming to assess Gata4, Mef2c, and Tbx5 occupancy. ChIP-seq in this system has proven extremely difficult for various reasons, such as limited starting material due to inefficient iCM generation. However, as we now know from scRNA-seq data, heterogeneity within reprogramming populations is likely to have hindered our ability to discern true GMT binding events. Notably, we completed our ChIP-seq experiments using antibodies against the endogenous forms of GMT rather than tagged proteins.

To assess the potential role of reprogramming factor binding on chromatin state, we plotted ATAC-seq signal in fibroblasts and day 2 iCMs, at sites bound by Gata4, Mef2c, or Tbx5 (Figure 4A). We then applied k-means clustering to define chromatin regions which remained statically open (purple) or were dynamically-open (green) (Figure 4A). Each of the three factors bound to both static and dynamically-open chromatin. To assess whether Gata4, Mef2c, or Tbx5 occupancy may regulate transcriptional changes by direct binding to promoters, we examined the distribution of distances to nearest TSS for each peak category. Regions bound by a reprogramming factor and associated with static open ATAC-seq signal were more often found within 2kb of a TSS compared with regions associated with dynamically-open ATAC-seq signal, suggesting that chromatin state changes directly affected by Gata4, Mef2c, or Tbx5 occupancy are more likely to occur at non-promoter proximal or distal regulatory regions (Figure 4B).

Next, we interrogated the possibility of cooperative binding between reprogramming factors, and its influence on chromatin state change. Many sites were bound by two or more factors and overlapped within domains that opened across our time course (Figure 4.1A). Notably, Gata4-bound regions associated with dynamic open chromatin state are more likely to also be bound by both Mef2c and Tbx5 than regions with static open chromatin state, suggesting cooperative binding between all reprogramming factors is more likely to remodel chromatin than Gata4 alone (Figure 4C). Similarly, regions bound by Tbx5 and associated with dynamic rather than static chromatin state are more likely to overlap with Gata4, Mef2c, or both (Figure 4B). Interestingly, Mef2C did not exhibit this phenomenon.

To further explore the relationship between individual reprogramming factor binding on chromatin accessibility changes, we plotted ATAC-seq signal at regions parsed based on single, double, or triple occupancy by reprogramming factors (Figure 4.2A). Regions bound by Gata4 or Tbx5 alone undergo no appreciable change in mean chromatin accessibility; however, at regions where the two peak sets overlap, average chromatin accessibility increases in day 2 iCMs compared with fibroblasts (Figure 4.2A-B). Mef2c is independently associated with opening chromatin. This association is also present when additional reprogramming factor binding sites overlap the Mef2c-bound region (Figure 4.2A-B).

Finally, we identified potential non-GMT cofactors by looking at *de novo* motifs enriched within ChIP-seq peaks, and matching these motifs to the closest known motif. As expected, top-ranked motifs correspond to families of reprogramming transcription factors targeted by ChIP. Interestingly, motifs ranking 2 and 3 differ between dynamic vs

static chromatin states, suggesting that different combinations of transcription factors may influence reprogramming through different regulatory strategies (Figure 4D).

In summary, here we present genomic binding locations for GMT at day 2 of direct cardiac reprogramming. By integrating this data with ATAC-seq data sets, we find that reprogramming factors are individually capable of binding at regions possessing different chromatin accessibility trends during reprogramming, with Mef2c most prominently associated to opening chromatin regions. In addition, GMT binding is more associated with promoter-distal regions at dynamically-opening loci compared with statically open loci. Furthermore, putative co-occupancy with another reprogramming factor is more common at dynamically rather than statically accessible sites for both Gata4 and Tbx5, suggesting co-operative binding by multiple reprogramming factors may in some cases be more effective than a single factor at affecting changes in chromatin state. These data will be essential in building models for how transcription factors interact with each other to execute direct cardiac reprogramming.

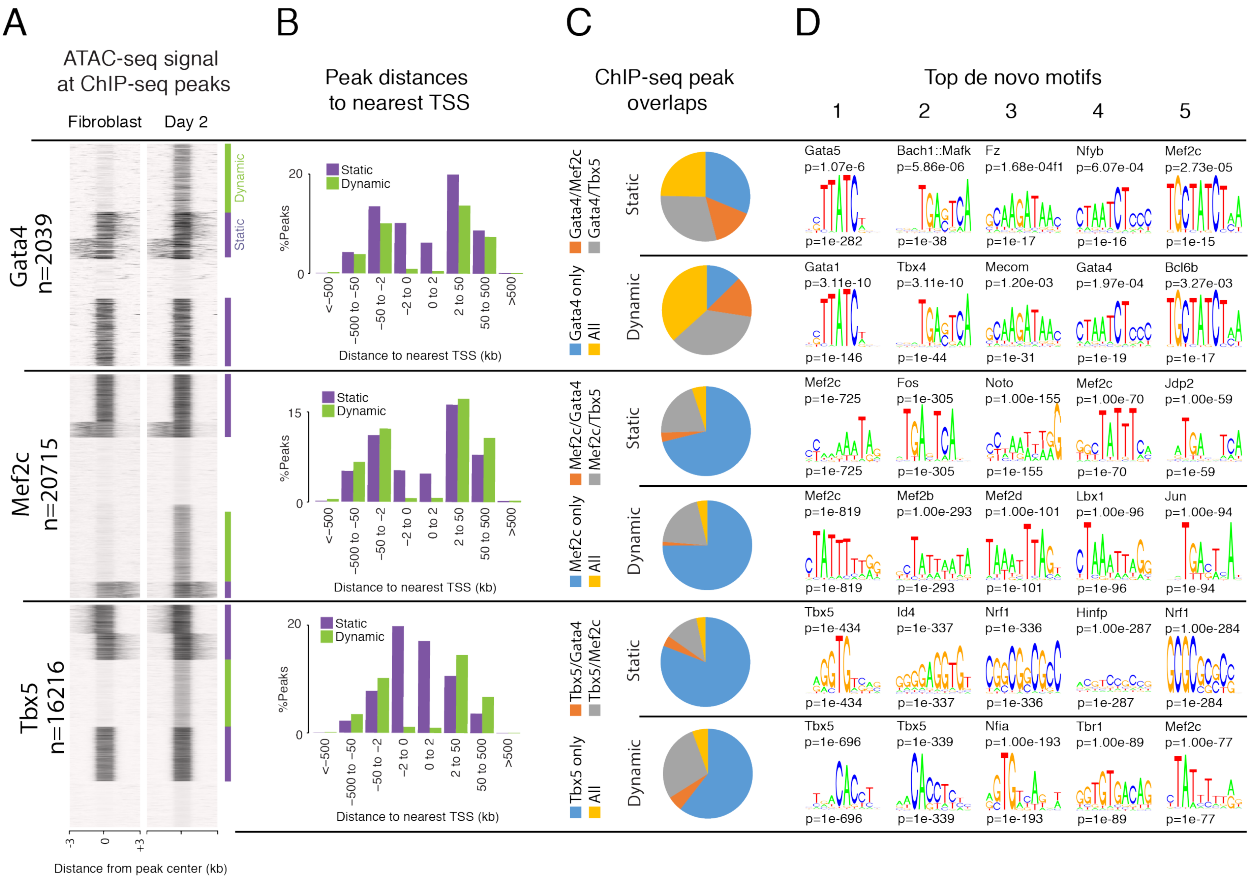


Figure 4. Chromatin accessibility dynamics associated with Gata4, Mef2c, and Tbx5 occupancy. (A) Heatmap showing ATAC-seq tag density at regions bound by reprogramming factors. We applied k-means clustering to define dynamic chromatin patterns of open “static” (blue), open “dynamic” (red), and closed. (B) Bar plots demonstrating ChIP-seq peak distributions relative to nearest TSS. Peak sets were parsed based on ATAC-seq dynamic patterns “static” (black) and “dynamic” (gray). (C) For each ChIP-seq peak set, pie charts show occurrence of putative reprogramming factor single, co-, or triple binding. (D) For each ChIP-seq peak set, we show sequence logos for the five top enriched *de novo* motifs compared with all open chromatin regions. p-value for *de novo* motifs is displayed below sequence logo. The best match known motif along with p-value is displayed above sequence logos.

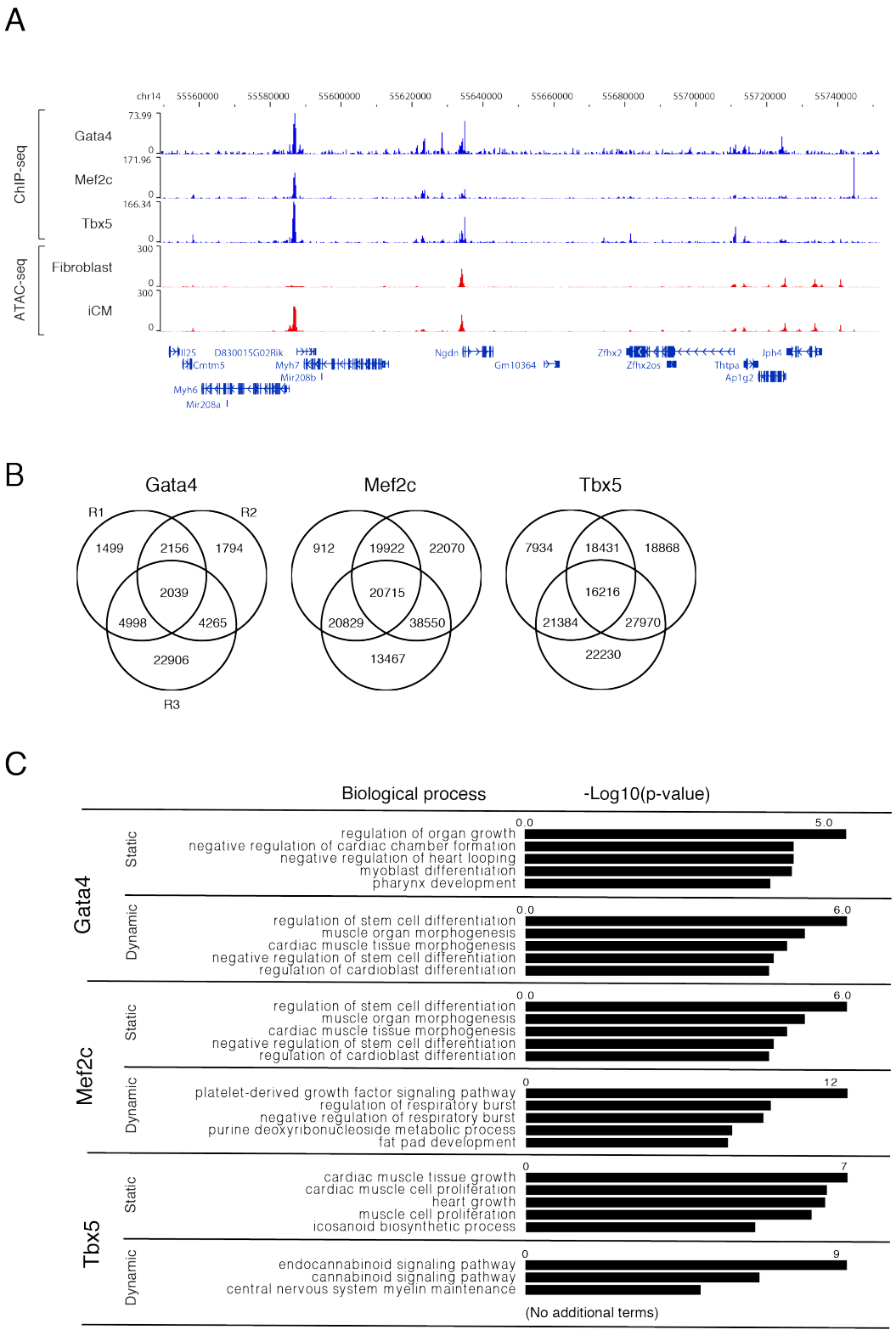


Figure 4.1. Biological processes associated with Gata4, Mef2c, Tbx5 occupancy. (A) Browser tracks illustrating reprogramming factor binding near cardiac loci. (B) Venn diagram shows intersection of peaks between three ChIP-seq replicates for Gata4,

Mef2c, and Tbx5. Fully replicate concordant peaks, shared between all three replicates, were used for subsequent analyses. (C) We parsed replicate concordant regions from reprogramming factor ChIP-seq, based on ATAC-seq dynamics (“static” or “dynamic”). Top 5 biological processes associated with these regions are listed. P-values are shown in bar graphs at right.

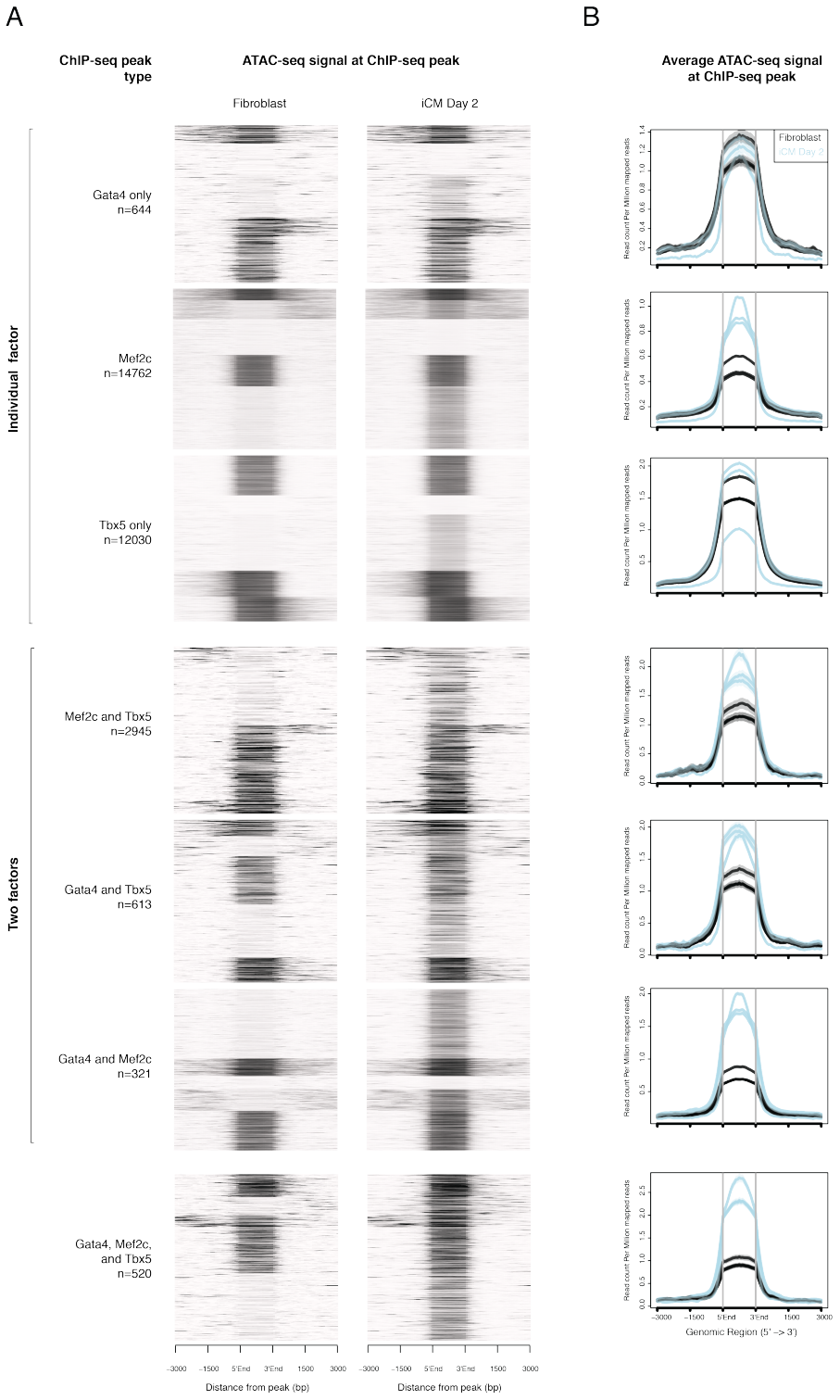


Figure 4.2. Regions bound by Gata4, Mef2c, and Tbx5, individually and in combination, display different chromatin accessibility trends. (A) Heatmaps show ATAC-seq signal from one representative replicate in fibroblast and day 2 iCM samples.

Each row is one replicate concordant peak identified from ChIP-seq for Gata4, Mef2c, Tbx5, or regions which intersect two or more of these peak sets. (B) Average ATAC-seq signal for three replicates over peak categories described in (A).

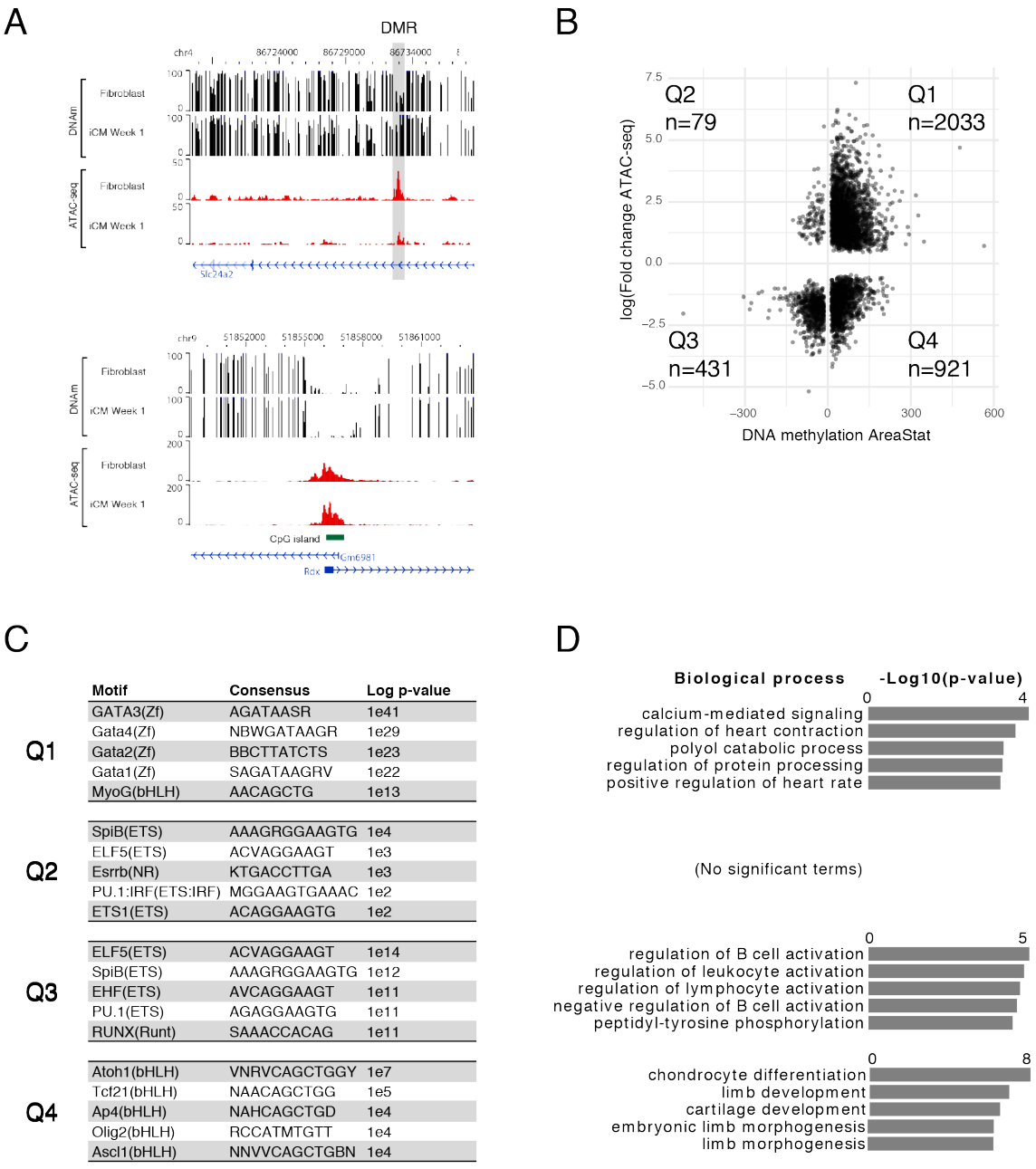
2.5 DNA methylation dynamics associated with reprogramming

DNA methylation is crucial to development and lineage fixation. To probe the role of DNA methylation in direct cardiac reprogramming, we performed whole-genome bisulfite-sequencing (WGBS), which detects genome-wide methylation at the single base resolution. We compared DNA methylation levels in fibroblasts vs week 1 iCMs generated using the GMTc protocol to define dynamically-methylated regions (DMRs) (Figure 5A). We identified 9,972 total DMRs, 6,707 hypermethylating and 2,960 hypomethylating, and further parsed them based on chromatin state information from ATAC-seq to define four DMR groups (Q1-Q4), associated with significant changes in both DNA methylation and ATAC-seq signal intensity (Figure 5B). Interestingly, while both high ATAC-seq signal and low DNA methylation levels are generally associated with increased transcription, these parameters coincide in the expected sense relatively rarely in our data sets, with the most represented group corresponding to simultaneously increasing DNA methylation and increasing ATAC-seq (Figure 5B).

To understand whether these different patterns may be the result of different groups of transcriptional regulators, we identified known transcription factor motifs enriched within each quadrant (Figure 5C). Interestingly, Gata family motifs are enriched in Q1, regions gaining both ATAC-seq signal, indicating opening chromatin, and DNA methylation. To further probe the potential biological importance of each differentially regulated region set, we defined top biological processes associated with each quadrant. Interestingly, along with Gata motif enrichment, regions in Q1 are also associated with cardiac-related terms (Figure 5D). This result suggests that direct

cardiac reprogramming by GMT involves DNA methylation at dynamically-open chromatin, with Gata4 being implicated in this unconventional pairing.

In summary, we identified dynamic DNA methylation at 9,972 genomic regions in week 1 iCMs. Regions gaining DNA methylation outnumbered those losing methylation by more than two-fold. Surprisingly, the majority of regions gaining DNA methylation also gain ATAC-seq signal. This unexpected subset is enriched for Gata4 motifs and is linked to cardiac function, suggesting relevance to direct cardiac reprogramming and a potentially novel regulatory mechanism in which a gain of DNA methylation (an epigenetic mark generally associated with inaccessible chromatin) is instead paired with increased chromatin accessibility.



C

	Motif	Consensus	Log p-value
Q1	GATA3(Zf)	AGATAASR	1e41
	Gata4(Zf)	NBWGATAAGR	1e29
	Gata2(Zf)	BBCTTATCTS	1e23
	Gata1(Zf)	SAGATAAGRV	1e22
	MyoG(bHLH)	AACAGCTG	1e13
Q2	SpiB(ETS)	AAAGRGGAAGTG	1e4
	ELF5(ETS)	ACVAGGAAGT	1e3
	Esrrb(NR)	KTGACCTTGA	1e3
	PU.1:IRF(ETS:IRF)	MGGAAGTGAAAC	1e2
	ETS1(ETS)	ACAGGAAGTG	1e2
Q3	ELF5(ETS)	ACVAGGAAGT	1e14
	SpiB(ETS)	AAAGRGGAAGTG	1e12
	EHF(ETS)	AVCAGGAAGT	1e11
	PU.1(ETS)	AGAGGAAGTG	1e11
	RUNX(Runt)	SAAACCACAG	1e11
Q4	Atoh1(bHLH)	VNRVCAGCTGGY	1e7
	Tcf21(bHLH)	NAACAGCTGG	1e5
	Ap4(bHLH)	NAHCAGCTGD	1e4
	Olig2(bHLH)	RCCATMTGTT	1e4
	Ascl1(bHLH)	NNVVCAGCTGBN	1e4

Figure 5. DNA methylation changes associated with reprogramming. (A) Genomic tracks showing WGBS (black) and ATAC-seq (red) data in fibroblasts and at week 1 of direct cardiac reprogramming at a hypermethylating DMR (upper) and at an open chromatin region containing a CpG island (lower). (B) Scatter plot showing methylation change (AreaStat) and log2 fold-change of ATAC-seq signal at DMRs containing significant ATAC-seq changes between week 1 iCMs and fibroblasts (Benjamini-Hochburg corrected $p < 0.05$). AreaStat is equal to the area of the t-statistic; equal to the sum of the t-statistics for individual methylation loci. (C) Top enriched *de novo* motifs per quadrant. (D) Bar plots illustrate top-ranked biological processes linked to quadrant regions.

CHAPTER 3
DISCUSSION

We interrogated the basis of direct cardiac reprogramming in an *in vitro* mouse system using FACS-enriched populations representing different stages of transdifferentiation. We provide extensive genomic data sets, and highlight several insights on the transcriptional and epigenetic dynamics that occur during this process.

3.1 Single-cell RNA-seq reveals mechanisms of reprogramming

Single-cell assays allow us to solve long-standing mysteries and clarify existing misinterpretation of bulk data sets. Similar to the OSKM cocktail used for iPSC reprogramming, the cardiac transcription factor combination of GMT was found through a combination of candidate approaches and systematic testing. Many advances since then have also employed this strategy, including the unbiased chemical screen that led to the discovery that TGFβ and Wnt inhibition greatly improved reprogramming yield and maturation (Mohamed, Stone et al. 2017).

Based on bulk RNA-seq analysis, one interpretation of how the GMTc protocol improved reprogramming was that it increased the speed of an iCM's progression through stages of reprogramming and pushed iCM maturation further than GMT alone (Figure 1.2). However, our transcriptional analysis of individual iCMs using scRNA-seq reveals that increased kinetics do not underly the observed improvements. Instead, we found that fibroblasts infected with GMT with or without chemical inhibition progress into largely the same cell populations and are capable of reaching the same terminal maturation point (Figure 1.1A-B). However, TGFβ and Wnt inhibition limited the proportion of fibroblasts that entered alternative reprogramming trajectories, thus a greater proportion became reporter-positive iCMs expressing cardiac markers, and reprogramming yield increased. Additionally, chemical inhibition limited the occurrence

of reporter-positive immune cells at later time points (Figure 1.1A-B). As a result, the average reporter-positive cell is closer to a cardiomyocyte when generated with GMTc, explaining the more mature, terminal iCM state suggested by bulk data alone.

Despite similarities to rapid transcriptional and chromatin remodeling seen in other systems, our results highlight differences between direct cardiac reprogramming and other reprogramming types. For example, during the transition from fibroblast to neuron the initial response to reprogramming factor expression is relatively homogenous, compared to the rapidly arising heterogeneity present by day 1 in the cardiac reprogramming system (Treutlein, Lee et al. 2016). During iPSC reprogramming, increasing cell division hastens the transition to a fully pluripotent state (Hanna, Saha et al. 2009). However, during human direct cardiac reprogramming, cell proliferation ceases within one division in cells that will become reporter-positive iCMs (Fu, Stone et al. 2013). Here, we demonstrate that failure to shut down expression of cell cycle genes is associated with intermediate reprogramming outcomes and an inability to progress to a more mature state (Figure 1.1C), offering insight into the mechanism through which proliferation hinders reprogramming.

3.2 Chromatin accessibility and DNA methylation dynamics highlight reprogramming complexity

One unexpected finding was the relationship between DNA methylation and ATAC-seq dynamics. Generally associated with stable silencing of gene expression and occurring at a final stage of cell fate change, we expected our most dynamically-methylated sites to be experiencing hypermethylation, correlated with the inverse trend in chromatin accessibility as measured by ATAC-seq. However, this was not the case;

the most prominent relationship between DNA methylation and chromatin accessibility dynamics shows a positive correlation between the direction of both methylation and accessibility change (Figure 5B, Q1). To test the possibility that this positive correlation was an artifact of the relatively small DMR length (mean DMR <200bp) obscuring broader changes in the vicinity of DMRs, we extended each DMR by 500bp and again compared to chromatin state changes (data not shown). Again, hypermethylation associated with gain of ATAC-seq signal was most representative of our DMR set. Interestingly, Q1 regions were enriched for Gata family motifs and cardiac-related biological processes, suggesting these may be regions undergoing opening by Gata4 and relevant to reprogramming (Figure 5C).

A variety of tolerance levels for methylated cytosine exists between transcription factor families, as well as within families. Indeed, many transcription factors crucial to development preferentially bind sequence motifs containing mCpG (Yin, Morgunova et al. 2017). Relevant to GMT reprogramming factors, some members of the Gata and T-box families display a preference for methylated motifs. MADS family motifs, including that of Mef2c, do not contain a CpG. Although the cognate Gata4 motif does not contain a CpG, motif enrichment within ChIP-seq peaks demonstrates that alternative motifs including Gata5, which displays a preference for methylated motif sequences, are more enriched at Gata4-bound regions than the canonical Gata4 motif (Figure 4D). Notably, the Gata motif most enriched in regions gaining both DNA methylation and ATAC-seq signal is Gata3, another Gata family member which has a documented preference for mCpG binding (Yin, Morgunova et al. 2017). This suggests that in the context of direct cardiac reprogramming, Gata4 may preferentially access alternative motif variants

which contain CpG methylation, and this phenomenon may underlie the unexpected correlation found between hypermethylation and chromatin opening during cardiac reprogramming.

Despite DNA methylation being generalized as a repressive epigenetic mark, this regulatory mechanism is more complex, with oxidation state and specific location of the methylated cytosine shaping its effect on transcription. For example, although methylation at enhancer regions is associated with transcriptional repression of targets, gene body methylation is associated with increased gene expression (Schubeler 2015). Additionally, the Qian group found that the methylation status of individual CpG's within cardiac promoters *Nppa* and *Myh6* were more predictive of mRNA levels than average methylation states of the entire promoter (Liu, Chen et al. 2016). Regarding oxidation state, 5-hydroxymethylation (5hmC) is associated with the progressive oxidation and removal of DNA methylation facilitated by the TET family of cytosine oxygenase enzymes (Ito, Shen et al. 2011). However, the 5hmC modification is not discernable from the non-oxidized form of cytosine based the WGBS approach. An alternative approach such as methylated DNA immunoprecipitation (MeDIP) would be necessary to determine whether the increases in DNA methylation we see during direct cardiac reprogramming are representative of 5hmC or 5mC (Weber, Davies et al. 2005).

Whether based on gain of 5hmC or 5mC, validation would be necessary to confirm the novel and unexpected relationship between DNAm and chromatin accessibility trends present in our data sets. One possibility is that these trends are not occurring in the same cells; the correlation between DNA methylation and chromatin state dynamics may simply be an artifact of bulk sampling of iCMs, reflecting an

average of heterogeneous populations undergoing different types of transitions. Given the available data sets, we are unable to directly test whether these trends co-occur in the same cells. To validate these results, we are exploring a protocol for “ATAC-BS-seq”, in which DNA methylation information is obtained only from accessible chromatin regions. ChIP-BS-seq is an approach in which chromatin immunoprecipitation of a feature of interest is followed by bisulfite treatment and subsequent library preparation to demonstrate direct correlation between these two modifications (Brinkman, Gu et al. 2012). Similarly, ATAC-BS-seq would directly quantify methylation status at accessible chromatin regions. Ideally, both DNA methylation and chromatin accessibility information would simultaneously be available from an individual cell, eliminating the need to interpret average values; however, this approach is not currently feasible.

3.3 Conclusions

We provide the first comprehensive genomic assessment of transcription factor binding, chromatin state and transcriptional changes, as well as thorough analysis of shifting identity of individual cells during direct cardiac reprogramming. Previously, our limited understanding of the molecular mechanism underlying cardiac reprogramming has significantly hindered our ability to make evidence-based improvements, and thus stunted its potential translational applications. Mechanistic insights provided here increase our understanding of direct cardiac reprogramming and will inform further improvements to this process.

3.4 Future directions

Currently, single-cell technologies allow us to discern cell identity-defining factors including transcription, transcription factor occupancy, chromatin accessibility, histone

modifications, and DNA methylation of individual cells. As a result, we are beginning to appreciate the heterogeneous outcomes resulting from all types of reprogramming examined thus far, and even discover previously undocumented, existing cell types (Luo, Keown et al. 2017, Villani, Satija et al. 2017). However, our ability to obtain multiple types of genomic information from the same cell remains limited.

In the near future, single-cell multi-omics approaches will allow us to apply multiple single-cell assays within the same cell. An assay called “scM&T-seq” has been used to obtain both transcription and DNA methylation from an individual cell and identify heterogeneous methylation patterns at regions regulating pluripotency (Angermueller, Clark et al. 2016). A “multi-omics” approach named “COOL-seq” has been applied to simultaneously assess genome-wide chromatin state, nucleosome positioning, DNA methylation, copy number variation, and ploidy from mouse preimplantation embryos and categorize genes based on the epigenetic heterogeneity of their promoters (Guo, Li et al. 2017). However, protocols that leverage more than two single-cell techniques simultaneously are not yet widespread. However, multi-omics is the next step in single-cell technologies, and once technical hurdles have been surmounted, this approach will undoubtedly be widely used in the future.

An ongoing challenge is designing analysis tools capable of keeping up with these new types of data. Recently, the Trapnell group developed a tool called CICERO which leverages scATAC-seq to reconstruct dynamic 3D chromatin architecture. Using this tool, they were able to identify the shifting genomic participants of “chromatin hubs” within MYOD-induced skeletal myoblast reprogramming (Pliner 2017). By designing programs like CICERO, scATAC-seq analysis becomes more accessible, and

examination of chromatin architecture, once limited by the resources necessary for complex assays such as Hi-C, becomes a basic step towards understanding a cell's identity.

Beyond the value in characterizing cell-fate changes, chromatin architecture data can be used to inform approaches for engineering cell fate, providing putative sites for targeted epigenomic remodeling using CRISPR-Cas9 technology. Recently, several groups have been able to achieve targeted epigenomic remodeling using a nuclease-dead dCas9 fused to the catalytic domains of epigenetic modifiers (Maeder, Angstman et al. 2013, Mendenhall, Williamson et al. 2013, Hilton, D'Ippolito et al. 2015, McDonald, Celik et al. 2016, Xu, Tao et al. 2016). These targeted epigenetic modifiers will be essential for parsing which epigenomic states associated with a given cell type are causative, and which are correlative.

Eventually, we may be able to catalogue every known transcriptional and epigenetic process occurring in a cell at any given time. With that, we will understand how transcription factors affect gene networks, how they interact with each other, and the molecular markers that define chromatin structure. Not only will we gain a more holistic view of the mechanisms behind cell fate transitions, we will gain a roadmap to manipulate the epigenome and cell identity.

CHAPTER 4
METHODS

Cell Culture

Direct cardiac reprogramming was performed on neonatal mouse cardiac fibroblasts as previously described, using only freshly prepared retroviruses (Ieda, Fu et al. 2010, Qian, Berry et al. 2013). ChIP-seq was performed using an immortalized neonatal cardiac fibroblast cell line, expressing cre-excisable T-antigen, developed by Palmer Yu in the Srivastava lab.

Sequencing Library Construction

Single-cell RNA sequencing (scRNA-seq)

Single cell RNA-seq libraries were prepared using the Chromium Single Cell 3' Reagent Kits v2 (PN-120236, PN-120237, PN-120262). Biological replicates were created for four time points: minus 1, plus 1, plus 7 and plus 14 days. All libraries were pooled and sequenced using the HiSeq 4000 to a read depth of at least 50,000 reads per cell.

Bulk RNA sequencing

Total RNA was isolated using the miRNeasy Micro Kit (Qiagen 217084). Bulk RNA-seq libraries were prepared with ovation RNA-seq system v2 kit (NuGEN). The RNA-seq libraries were analyzed by Bioanalyzer and quantified by QPCR (KAPA). Samples were sequenced at 100PE on the Illumina HiSeq 2500 at the Harvard FAS core.

Assay for Transposase-Accessible Chromatin with Deep Sequencing (ATAC-seq)

We prepared iCM, single factor, and fibroblast and neonatal CM samples for ATAC-seq based on the protocol previously described (Buenrostro, Giresi et al. 2013). Aliquots of 10,000 - 50,000 cells were spun down (310 RCF for 3 minutes) and washed with 200 µL chilled PBS. Samples were lysed with 200 µL chilled lysis buffer (20mM Tris-HCl (pH 8.0), 85mM KCl, 0.5% NP-40) and spun down at 500 RCF for 5 minutes. Nuclear pellets were then each transposed with 25 µL Tagment DNA Buffer, 2.5 µL Tagment DNA Enzyme (Nextera Sample prep Kit from Illumina, cat # FC-121-1030), and 22.5 µL Nuclease-Free H_2O. The samples were then incubated at 37°C for 30 minutes and stored at −20°C. Transposed samples were purified using the QIAGEN MinElute Reaction Cleanup Kit (cat #28204). Samples were then amplified using 25 µL NEBNext High Fidelity 2x PCR Master Mix, 1.25 µM Nextera custom primer (common to all samples), 1.25 µM Nextera custom primers with unique barcodes, and Nuclease-Free H_2O. We amplified samples using the following PCR conditions: 72°C for 5 minutes; 98°C for 30 seconds; and thermocycling at 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 1 minute. Half of each sample was amplified for 12 cycles, MinElute purified and assessed by bioanalyzer for library quality. Samples concentration was quantified by Qbit before pooling. Samples were sequenced at 100PE on the Illumina HiSeq 2500 at either the Harvard FAS core or the UCSF CAT core.

Whole genome bisulfite sequencing (WGBS)

For WGBS sequencing libraries, genomic DNA was isolated using the Qiagen DNeasy Blood and Tissue kit and fragmented using a Covaris S2 Sonicator. Libraries were

prepared with 10ug gDNA using the Ovation® Ultralow Methyl Seq DR Multiplex System 1-8, 9-16 (NuGen 0335 and 0336). Libraries were pooled and sequenced at 100PE on a HiSeq 2000 in standard run mode to a read depth of >180M mapped reads.

Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq)

Cells (10^6 per ChIP) were crosslinked in 1% formaldehyde in suspension at room temperature for 10 minutes with gentle rotation. Crosslinking was quenched by addition of glycine (final 125mM) followed by incubation at room temperature for 5 minutes with gentle rotation. Cell pellets were lysed in buffer [20mM Tris-HCl pH 8, 85mM KCl, 0.5% NP-40, protease inhibitors] for 10 minutes on a rotator at 4°C. Nuclei were isolated by centrifugation (2,500 x g, 5 minutes, 4°C), resuspended in nuclei lysis buffer [50mM Tris-HCl pH 8, 10mM EDTA pH 8, 1% SDS, protease inhibitors] and incubated on a rotator for 30 minutes at 4°C. Chromatin was sheared using a Covaris S2 sonicator for 15 minutes (60-second cycles, 5% duty cycle, 200 cycles/burst, intensity = 5) until DNA was in the 200 – 700 base-pair range. Chromatin was diluted 5-fold in ChIP dilution buffer [0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8, 167mM NaCl, protease inhibitors] and incubated with antibody (2ug/million cells) at 4°C overnight under rotation. Antibody-protein complexes were immunoprecipitated using Pierce Protein A/G magnetic beads at 4°C for 2 hours under rotation. Beads were washed 5 times (2 minute washes under rotation) with cold RIPA buffer [50mM HEPES-KOH pH 7.5, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate] followed by 1 wash in cold final wash buffer [1xTE, 50mM NaCl]. Immunoprecipitated chromatin was eluted at 65°C with agitation for 30 minutes in elution buffer [50mM Tris-HCl pH 8.0,

10mM EDTA, 1% SDS]. High salt buffer [250mM Tris-Hcl pH 7.5, 32.5mM EDTA pH 8, 1.25M NaCl] and Proteinase K were added and crosslinks were reversed overnight at 65°C. Samples were treated with RNase A and DNA was purified with Agencourt AMPre XP beads. Fragmented ChIP and control (whole-cell extract) DNA was end-repaired, 5' phosphorylated and dA-tailed with NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645). Samples were ligated to adaptor oligos for multiplex sequencing (NEB E7335), PCR amplified and sequenced on an Illumina NextSeq500.

Sequencing data processing and analyses

Single cell RNA-seq analysis

The 10X Genomics Cell Ranger pipeline was used to demultiplex raw data, align reads, count transcripts and aggregate multiple samples and timepoints. The R package "Seurat" (Butler 2017) was used for all downstream analyses. Cells that met unique molecular index (UMI) and gene thresholds were included. The data was log normalized and the most variable genes were identified by dispersion. Technical variation due to genes and UMIs detected in each cell was regressed out. Statistically significant principal components based on the most variable genes were then calculated using the Jackstraw method and then utilized to identify clusters. Differential expression between the clusters was calculated using the "LRT test based on zero-inflated data" test option.

Bulk RNA-seq analysis

Trimming of known adapters and low-quality regions of reads was performed using Fastq-mcf (http://code.google.com/p/ea-utils). Sequence quality control was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and RSeQC (Wang, Wang et al. 2012). Alignment of the provided samples to the mm9 reference genome was performed using Bowtie 2.2.4 (Langmead and Salzberg 2012). Reads were assigned to genes using "featureCounts" (Liao, Smyth et al. 2014), part of the Subread suite (http://subread.sourceforge.net/). Gene-level counts were arrived at using Ensembl gene annotation, in GTF format. We calculated differential expression P-values using edgeR, an R package available through Bioconductor (Robinson, McCarthy et al. 2010). We first filtered out genes where there were not at least two samples with at least 5 (raw) reads. Once these genes are removed, we recalculate the counts per million for each gene (CPM) and filter out any genes with a CPM above 20,000. After excluding these genes, we re-normalize the remaining ones using calcNormFactors (TMM) ("weighted trimmed mean of M-values") in edgeR (Robinson and Oshlack 2010). Calculation of P-values is performed in edgeR for the differential expression between samples. EdgeR uses a negative binomial distribution as a model for expected gene expression. Finally we use the built-in R function "p.adjust" to calculate the FDR (false discovery rate) for each P-value, using the Benjamini-Hochberg method (Benjamini and Hochberg 1995). We used GO-Elite to generate biological ontology terms (Zambon, Gaj et al. 2012).

ATAC-seq analysis

Reads were aligned to the mm9 genomic assembly using bowtie2 with options: -X 600 --no-mixed --no-discordant. Duplicate reads were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard). Peaks were called using macs2 callpeak with options: -p 0.1 --nomodel --shift 100 --extsize 200 -B --SPMR --call-summits. Peaks concordant between at least two of three replicates were considered for further analysis. Data for Figure 3 heatmap was clustered using HOPACH method (Van der Laan 2003). We used ngs.plot to display dynamic ATAC-seq signal at GMT-bound regions (Shen, Shao et al. 2014). We associated biological processes to these regions using GREAT (McLean, Bristor et al. 2010). Motif enrichment analysis was completed using HOMER (Heinz, Benner et al. 2010).

ChIP-seq analysis

Trimming of known adapters and low-quality regions of reads was performed using Fastq-mcf (http://code.google.com/p/ea-utils). Sequence quality control was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and RSeQC (Wang, Wang et al. 2012). Alignment of the provided samples to the mm9 reference genome was performed using Bowtie 2.2.4 (Langmead and Salzberg 2012). Peaks were called using GEM (Guo, Mahony et al. 2012). To ensure a high degree of stringency, only fully concordant peaks called in all three replicates with an overlap of >50% were used for subsequent analyses (Figure 4.1B). Peaks with an overlap of at least 1bp from different transcription factor ChIP-seq experiments were considered candidate cobinding events. Region intersects were found using BEDTools (Quinlan

and Hall 2010). Motif enrichment analysis was completed using HOMER (Heinz, Benner et al. 2010). Known motifs and sequence logos were generated from *de novo motifs* matched to the JASPAR CORE non-redundant vertebrate database (Mathelier, Fornes et al. 2016) using Tomom from the MEME suite (Bailey, Boden et al. 2009).

WGBS analysis

Reads were trimmed using Fastq-mcf (http://code.google.com/p/ea-utils) and aligned to the mm9 genomic assembly. Sequencing quality was assessed with FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and duplicate reads were removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard). Methylation calls were made with Bismark bismark_methylation_extractor using options: -p -no_overlap. DMRs were identified using Bsseq (Hansen, Langmead et al. 2012). We required minimum CpG content of >=3 and >=0.1 absolute difference in methylation for all DMRs. Motif enrichment analysis was completed using HOMER (Heinz, Benner et al. 2010).

REFERENCES

Abad, M., H. Hashimoto, H. Zhou, M. G. Morales, B. Chen, R. Bassel-Duby and E. N. Olson (2017). "Notch Inhibition Enhances Cardiac Reprogramming by Increasing MEF2C Transcriptional Activity." Stem Cell Reports **8**(3): 548-560.

Aldiri, I., B. Xu, L. Wang, X. Chen, D. Hiler, L. Griffiths, M. Valentine, A. Shirinifard, S. Thiagarajan, A. Sablauer, M. E. Barabas, J. Zhang, D. Johnson, S. Frase, X. Zhou, J. Easton, J. Zhang, E. R. Mardis, R. K. Wilson, J. R. Downing, M. A. Dyer and P. St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome (2017). "The Dynamic Epigenetic Landscape of the Retina During Development, Reprogramming, and Tumorigenesis." Neuron **94**(3): 550-568 e510.

Angermueller, C., S. J. Clark, H. J. Lee, I. C. Macaulay, M. J. Teng, T. X. Hu, F. Krueger, S. Smallwood, C. P. Ponting, T. Voet, G. Kelsey, O. Stegle and W. Reik (2016). "Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity." Nat Methods **13**(3): 229-232.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble (2009). "MEME SUITE: tools for motif discovery and searching." Nucleic Acids Res **37**(Web Server issue): W202-208.

Brinkman, A. B., H. Gu, S. J. Bartels, Y. Zhang, F. Matarese, F. Simmer, H. Marks, C. Bock, A. Gnirke, A. Meissner and H. G. Stunnenberg (2012). "Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk." Genome Res **22**(6): 1128-1138.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Butler, A., R. Satija (2017). "Integrated analysis of single cell transcriptomic data across conditions, technologies, and species." bioRxiv.

Chong, J. J., X. Yang, C. W. Don, E. Minami, Y. W. Liu, J. J. Weyers, W. M. Mahoney, B. Van Biber, S. M. Cook, N. J. Palpant, J. A. Gantz, J. A. Fugate, V. Muskheli, G. M. Gough, K. W. Vogel, C. A. Astley, C. E. Hotchkiss, A. Baldessari, L. Pabon, H. Reinecke, E. A. Gill, V. Nelson, H. P. Kiem, M. A. Laflamme and C. E. Murry (2014). "Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts." Nature **510**(7504): 273-277.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Davis, R. L., H. Weintraub and A. B. Lassar (1987). "Expression of a single transfected cDNA converts fibroblasts to myoblasts." Cell **51**(6): 987-1000.

Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis and B. E. Bernstein (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.

Feng, R., S. C. Desbordes, H. Xie, E. S. Tillo, F. Pixley, E. R. Stanley and T. Graf (2008). "PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells." Proc Natl Acad Sci U S A **105**(16): 6057-6062.

Fields, P. A., V. Ramani, G. Bonora, G. Yardimci, A. Bertero, H. Reinecke, L. Pabon, W. S. Noble, J. Shendure, C. E. Murry (2017). "Dynamic reorganization of nuclear architecture during human cardiogenesis." bioRxiv.

Fu, J. D., N. R. Stone, L. Liu, C. I. Spencer, L. Qian, Y. Hayashi, P. Delgado-Olguin, S. Ding, B. G. Bruneau and D. Srivastava (2013). "Direct Reprogramming of Human Fibroblasts toward a Cardiomyocyte-like State." Stem Cell Reports **1**(3): 235-247.

Garcia-Castro, J. and I. Singec (2017). "Prospects of Pluripotent and Adult Stem Cells for Rare Diseases." Adv Exp Med Biol **1031**: 371-386.

Gifford, C. A., M. J. Ziller, H. Gu, C. Trapnell, J. Donaghey, A. Tsankov, A. K. Shalek, D. R. Kelley, A. A. Shishkin, R. Issner, X. Zhang, M. Coyne, J. L. Fostel, L. Holmes, J. Meldrim, M. Guttman, C. Epstein, H. Park, O. Kohlbacher, J. Rinn, A. Gnirke, E. S. Lander, B. E. Bernstein and A. Meissner (2013). "Transcriptional and epigenetic dynamics during specification of human embryonic stem cells." Cell **153**(5): 1149-1163.

Gilsbach, R., S. Preissl, B. A. Gruning, T. Schnick, L. Burger, V. Benes, A. Wurch, U. Bonisch, S. Gunther, R. Backofen, B. K. Fleischmann, D. Schubeler and L. Hein (2014). "Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease." Nat Commun **5**: 5288.

Guenther, M. G., S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young (2007). "A chromatin landmark and transcription initiation at most promoters in human cells." Cell **130**(1): 77-88.

Guo, F., L. Li, J. Li, X. Wu, B. Hu, P. Zhu, L. Wen and F. Tang (2017). "Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells." Cell Res **27**(8): 967-988.

Guo, Y., S. Mahony and D. K. Gifford (2012). "High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints." PLoS Comput Biol **8**(8): e1002638.

Gurdon, J. B. (1962). "The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles." J Embryol Exp Morphol **10**: 622-640.

Hanna, J., K. Saha, B. Pando, J. van Zon, C. J. Lengner, M. P. Creyghton, A. van Oudenaarden and R. Jaenisch (2009). "Direct cell reprogramming is a stochastic process amenable to acceleration." Nature **462**(7273): 595-601.

Hansen, K. D., B. Langmead and R. A. Irizarry (2012). "BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions." Genome Biol **13**(10): R83.

Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford and B. Ren (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-318.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.

Hilton, I. B., A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy and C. A. Gersbach (2015). "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers." Nat Biotechnol **33**(5): 510-517.

Hirai, H. and N. Kikyo (2014). "Inhibitors of suppressive histone modification promote direct reprogramming of fibroblasts to cardiomyocyte-like cells." Cardiovasc Res **102**(1): 188-190.

Hong, S. G., T. Winkler, C. Wu, V. Guo, S. Pittaluga, A. Nicolae, R. E. Donahue, M. E. Metzger, S. D. Price, N. Uchida, S. A. Kuznetsov, T. Kilts, L. Li, P. G. Robey and C. E. Dunbar (2014). "Path to the clinic: assessment of iPSC-based cell therapies in vivo in a nonhuman primate model." Cell Rep **7**(4): 1298-1309.

Huang, P., Z. He, S. Ji, H. Sun, D. Xiang, C. Liu, Y. Hu, X. Wang and L. Hui (2011). "Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors." Nature **475**(7356): 386-389.

Ieda, M., J. D. Fu, P. Delgado-Olguin, V. Vedantham, Y. Hayashi, B. G. Bruneau and D. Srivastava (2010). "Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors." Cell **142**(3): 375-386.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." Science **333**(6047): 1300-1303.

Kajimura, S., P. Seale, K. Kubota, E. Lunsford, J. V. Frangioni, S. P. Gygi and B. M. Spiegelman (2009). "Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex." Nature **460**(7259): 1154-1158.

Kimbrel, E. A. and R. Lanza (2015). "Current status of pluripotent stem cells: moving the first therapies to the clinic." Nat Rev Drug Discov **14**(10): 681-692.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." Bioinformatics **30**(7): 923-930.

Liu, Z., O. Chen, M. Zheng, L. Wang, Y. Zhou, C. Yin, J. Liu and L. Qian (2016). "Re-patterning of H3K27me3, H3K4me3 and DNA methylation during fibroblast conversion into induced cardiomyocytes." Stem Cell Res **16**(2): 507-518.

Luo, C., C. L. Keown, L. Kurihara, J. Zhou, Y. He, J. Li, R. Castanon, J. Lucero, J. R. Nery, J. P. Sandoval, B. Bui, T. J. Sejnowski, T. T. Harkins, E. A. Mukamel, M. M. Behrens and J. R. Ecker (2017). "Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex." Science **357**(6351): 600-604.

Maeder, M. L., J. F. Angstman, M. E. Richardson, S. J. Linder, V. M. Cascio, S. Q. Tsai, Q. H. Ho, J. D. Sander, D. Reyon, B. E. Bernstein, J. F. Costello, M. F. Wilkinson and J. K. Joung (2013). "Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins." Nat Biotechnol **31**(12): 1137-1142.

Mathelier, A., O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin and W. W. Wasserman (2016). "JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles." Nucleic Acids Res **44**(D1): D110-115.

McDonald, J. I., H. Celik, L. E. Rois, G. Fishberger, T. Fowler, R. Rees, A. Kramer, A. Martens, J. R. Edwards and G. A. Challen (2016). "Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation." Biol Open **5**(6): 866-874.

McLean, C. Y., D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger and G. Bejerano (2010). "GREAT improves functional interpretation of cis-regulatory regions." Nat Biotechnol **28**(5): 495-501.

Mendenhall, E. M., K. E. Williamson, D. Reyon, J. Y. Zou, O. Ram, J. K. Joung and B. E. Bernstein (2013). "Locus-specific editing of histone modifications at endogenous enhancers." Nat Biotechnol **31**(12): 1133-1136.

Mohamed, T. M., N. R. Stone, E. C. Berry, E. Radzinsky, Y. Huang, K. Pratt, Y. S. Ang, P. Yu, H. Wang, S. Tang, S. Magnitsky, S. Ding, K. N. Ivey and D. Srivastava (2017). "Chemical Enhancement of In Vitro and In Vivo Direct Cardiac Reprogramming." Circulation **135**(10): 978-995.

Nam, Y. J., K. Song, X. Luo, E. Daniel, K. Lambeth, K. West, J. A. Hill, J. M. DiMaio, L. A. Baker, R. Bassel-Duby and E. N. Olson (2013). "Reprogramming of human fibroblasts toward a cardiac fate." Proc Natl Acad Sci U S A **110**(14): 5588-5593.

Papp, B. and K. Plath (2013). "Epigenetics of reprogramming to induced pluripotency." Cell **152**(6): 1324-1343.

Pliner, H., J. Packer, J. McFaline-Figueroa, D. Cusanovich, R. Daza, S. Srivatsan, X. Qiu, D. Jackson, A. Minkina, A. Ade y, F. Steemers, J. Shendure, C. Trapnell (2017). "Chromatin accessibility dynamics of myogenesis at single cell resolution." bioRxiv.

Polo, J. M., E. Anderssen, R. M. Walsh, B. A. Schwarz, C. M. Nefzger, S. M. Lim, M. Borkent, E. Apostolou, S. Alaei, J. Cloutier, O. Bar-Nur, S. Cheloufi, M. Stadtfeld, M. E. Figueroa, D. Robinton, S. Natesan, A. Melnick, J. Zhu, S. Ramaswamy and K. Hochedlinger (2012). "A molecular roadmap of reprogramming somatic cells into iPS cells." Cell **151**(7): 1617-1632.

Qian, L., E. C. Berry, J. D. Fu, M. Ieda and D. Srivastava (2013). "Reprogramming of mouse fibroblasts into cardiomyocyte-like cells in vitro." Nat Protoc **8**(6): 1204-1215.

Qian, L., Y. Huang, C. I. Spencer, A. Foley, V. Vedantham, L. Liu, S. J. Conway, J. D. Fu and D. Srivastava (2012). "In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes." Nature **485**(7400): 593-598.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.

Robinson, M. D. and A. Oshlack (2010). "A scaling normalization method for differential expression analysis of RNA-seq data." Genome Biol **11**(3): R25.

Schubeler, D. (2015). "Function and information content of DNA methylation." Nature **517**(7534): 321-326.

Sekiya, S. and A. Suzuki (2011). "Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors." Nature **475**(7356): 390-393.

Shen, L., N. Shao, X. Liu and E. Nestler (2014). "ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases." BMC Genomics **15**: 284.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.

Treutlein, B., Q. Y. Lee, J. G. Camp, M. Mall, W. Koh, S. A. Shariati, S. Sim, N. F. Neff, J. M. Skotheim, M. Wernig and S. R. Quake (2016). "Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq." Nature **534**(7607): 391-395.

Van der Laan, M. J., K. S. Pollard (2003). "Hybrid clustering of gene expression data with visualization and the bootstrap." Journal of Statistical Planning and Inference **117**: 275-303.

Vierbuchen, T., A. Ostermeier, Z. P. Pang, Y. Kokubu, T. C. Sudhof and M. Wernig (2010). "Direct conversion of fibroblasts to functional neurons by defined factors." Nature **463**(7284): 1035-1041.

Vierbuchen, T. and M. Wernig (2012). "Molecular roadblocks for cellular reprogramming." Mol Cell **47**(6): 827-838.

Villani, A. C., R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P. L. De Jager, O. Rozenblatt-Rosen, A. A. Lane, M. Haniffa, A. Regev and N. Hacohen (2017). "Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors." Science **356**(6335).

Wamstad, J. A., J. M. Alexander, R. M. Truty, A. Shrikumar, F. Li, K. E. Eilertson, H. Ding, J. N. Wylie, A. R. Pico, J. A. Capra, G. Erwin, S. J. Kattman, G. M. Keller, D. Srivastava, S. S. Levine, K. S. Pollard, A. K. Holloway, L. A. Boyer and B. G. Bruneau (2012). "Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage." Cell **151**(1): 206-220.

Wang, L., Z. Liu, C. Yin, H. Asfour, O. Chen, Y. Li, N. Bursac, J. Liu and L. Qian (2015). "Stoichiometry of Gata4, Mef2c, and Tbx5 influences the efficiency and quality of induced cardiac myocyte reprogramming." Circ Res **116**(2): 237-244.

Wang, L., Z. Liu, C. Yin, Y. Zhou, J. Liu and L. Qian (2015). "Improved Generation of Induced Cardiomyocytes Using a Polycistronic Construct Expressing Optimal Ratio of Gata4, Mef2c and Tbx5." J Vis Exp(105).

Wang, L., S. Wang and W. Li (2012). "RSeQC: quality control of RNA-seq experiments." Bioinformatics **28**(16): 2184-2185.

Wapinski, O. L., Q. Y. Lee, A. C. Chen, R. Li, M. R. Corces, C. E. Ang, B. Treutlein, C. Xiang, V. Baubet, F. P. Suchy, V. Sankar, S. Sim, S. R. Quake, N. Dahmane, M. Wernig and H. Y. Chang (2017). "Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons." Cell Rep **20**(13): 3236-3247.

Wapinski, O. L., T. Vierbuchen, K. Qu, Q. Y. Lee, S. Chanda, D. R. Fuentes, P. G. Giresi, Y. H. Ng, S. Marro, N. F. Neff, D. Drechsel, B. Martynoga, D. S. Castro, A. E. Webb, T. C. Sudhof, A. Brunet, F. Guillemot, H. Y. Chang and M. Wernig (2013). "Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons." Cell **155**(3): 621-635.

Weber, M., J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam and D. Schubeler (2005). "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." Nat Genet **37**(8): 853-862.

Weintraub, H., S. J. Tapscott, R. L. Davis, M. J. Thayer, M. A. Adam, A. B. Lassar and A. D. Miller (1989). "Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD." Proc Natl Acad Sci U S A **86**(14): 5434-5438.

Xu, X., Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G. L. Xu and R. Hu (2016). "A CRISPR-based approach for targeted DNA demethylation." Cell Discov **2**: 16009.

Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schubeler, C. Vinson and J. Taipale (2017). "Impact of cytosine methylation on DNA binding specificities of human transcription factors." Science **356**(6337).

Zambon, A. C., S. Gaj, I. Ho, K. Hanspers, K. Vranizan, C. T. Evelo, B. R. Conklin, A. R. Pico and N. Salomonis (2012). "GO-Elite: a flexible solution for pathway and ontology over-representation." Bioinformatics **28**(16): 2209-2210.

Zhou, H., M. E. Dickson, M. S. Kim, R. Bassel-Duby and E. N. Olson (2015). "Akt1/protein kinase B enhances transcriptional reprogramming of fibroblasts to functional cardiomyocytes." Proc Natl Acad Sci U S A **112**(38): 11864-11869.

Zhou, Y., L. Wang, H. R. Vaseghi, Z. Liu, R. Lu, S. Alimohamadi, C. Yin, J. D. Fu, G. G. Wang, J. Liu and L. Qian (2016). "Bmi1 Is a Key Epigenetic Barrier to Direct Cardiac Reprogramming." Cell Stem Cell **18**(3): 382-395.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

12/19/2017

Date