Context-Specific Transcription Factor Functions Regulate Epigenomic and Transcriptional Dynamics during Cardiac Reprogramming

DOI: 10.1016/j.stem.2019.06.012

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Cell Stem Cell

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Unique Transcription Factor Functions Regulate Epigenomic and Transcriptional Dynamics During Cardiac Reprogramming

Nicole R. Stone\textsuperscript{1,2,3}, Casey A. Gifford\textsuperscript{1,2,4}, Reuben Thomas\textsuperscript{2}, Karishma J. B. Pratt\textsuperscript{2}, Kaitlen Samse-Knapp\textsuperscript{2}, Tamer M. A. Mohamed\textsuperscript{2,4}, Ethan M. Radzinsky\textsuperscript{2}, Amelia Schricker\textsuperscript{2}, Lin Ye\textsuperscript{2}, Pengzhi Yu\textsuperscript{2,4}, Joke G. van Bemmel\textsuperscript{2}, Kathryn N. Ivey\textsuperscript{2,3,4}, Katherine S. Pollard\textsuperscript{2,5,6,7}, and Deepak Srivastava\textsuperscript{2,3,4,7,8}

\textsuperscript{1}co-first authors
\textsuperscript{2}Gladstone Institutes, San Francisco, CA, 94158, USA
\textsuperscript{3}Departments of Pediatrics and Biochemistry & Biophysics, University of California, San Francisco, CA, 94143, USA
\textsuperscript{4}Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA, 94158, USA
\textsuperscript{5}Department of Epidemiology & Biostatistics, University of California, San Francisco, CA, 94143, USA
\textsuperscript{6}Chan-Zuckerberg Biohub, San Francisco, CA, 94158, USA
\textsuperscript{7}Corresponding authors
\textsuperscript{8}Lead contact

Correspondence to:
Deepak Srivastava (dsvivastava@gladstone.ucsf.edu)
Gladstone Institutes
1650 Owens St.
San Francisco, CA 94158
Phone: 415-734-2716
Fax: 415-355-0141

Katherine S. Pollard (katherine.pollard@gladstone.ucsf.edu)
Gladstone Institutes
1650 Owens St.
San Francisco, CA 94158
Phone: 415-734-2711
Fax: 415-355-0960
Key Words: Reprogramming; cardiomyocyte; cardiac fibroblast; transcription factor; ATAC-seq; single cell RNA-seq; ChIP-seq
SUMMARY

Ectopic expression of combinations of transcription factors (TF) can drive direct lineage conversion, thereby reprogramming a somatic cell's identity. To determine the molecular mechanisms by which Gata4, Mef2c, and Tbx5 (GMT) induce conversion from a cardiac fibroblast toward an induced cardiomyocyte, we performed comprehensive transcriptomic, DNA-occupancy, and epigenomic interrogation of the reprogramming process. Integration of these data sets identified new TFs involved in the reprogramming process and revealed context-specific roles for GMT, including the ability of Mef2c and Tbx5 to independently promote chromatin remodeling at previously inaccessible sites. We also find evidence for cooperative facilitation and refinement of each TF’s binding profile in a combinatorial setting. A reporter assay employing newly defined regulatory elements confirmed binding of a single TF can be sufficient for activation, suggesting co-binding events do not necessarily reflect synergy. These results shed light on fundamental mechanisms by which TFs direct lineage conversion.
INTRODUCTION

Somatic cellular identity is established by complex gene regulatory networks during embryonic development. Knowledge of these networks has been exploited to devise combinations of transcription factors that facilitate direct reprogramming of somatic cells from one lineage to another without progression through an intermediate pluripotent state (Feng et al., 2008; Ieda et al., 2010; Rackham et al., 2016; Wernig et al., 2008). However, the precise mechanisms by which various combinations lead to cellular specificity is only beginning to be understood (Treutlein et al., 2016; Wapinski et al., 2017).

Direct reprogramming of cardiac fibroblasts to cardiomyocytes has been achieved in a variety of ways, including ectopic expression of cardiac-enriched transcription factors (Fu et al., 2013; Ieda et al., 2010; Nam et al., 2013; Qian et al., 2012; Song et al., 2012). Ectopic expression of Gata4, Mef2c, and Tbx5 (GMT) is sufficient to alter the fibroblast epigenome and promote expression of genes associated with cardiomyocytes while simultaneously repressing the fibroblast gene program (Ieda et al., 2010; Liu et al., 2017; Zhou et al., 2016). Perturbation of epigenetic remodelers and ectopic expression of additional transcription factors such as Hand2 and Myocd, or microRNAs, were found to influence reprogramming as well, with addition of Hand2 resulting in a greater portion of pacemaker-like cells (Addis et al., 2013; Christoforou et al., 2013; Jayawardena et al., 2015; Nam et al., 2014; Protze et al., 2012; Zhou et al., 2016). Concomitant chemical inhibition of TGFβ and Wnt signaling resulted in improved reprogramming both in vitro and in vivo (Ifkovits et al., 2014; Mohamed et al., 2017). Introduction of reprogramming factors directly into the heart after damage by a gene therapy approach resulted in significantly improved cardiac function, suggesting potential therapeutic benefits of in vivo cardiac reprogramming for regenerative medicine (Jayawardena et al., 2015; Qian et al., 2012; Song et al., 2012).

Cardiac reprogramming factors, like other reprogramming factors, are typically tissue-enriched rather than tissue-specific, yet somehow still induce a unique fate switch. Gata4, Mef2c, and Tbx5 each have essential functions in a wide range of tissues during embryonic development, and their deletion individually leads to embryonic lethality and gross malformation of various organ systems, including the
developing cardiovascular system (Bruneau et al., 2001; Lin et al., 1997; Molkentin et al., 1997). Gata4 interacts with Nkx2-5 and Tbx5 to promote cardiovascular development, but it also cooperates with Foxa2 during endoderm development to promote expression of a transcriptional network required for foregut development (Holtzinger and Evans, 2005). Similarly, beyond cardiogenesis Tbx5 is also required for limb development (Agarwal et al., 2003; Ahn et al., 2002), while Mef2c is also essential for neural development (Li et al., 2008; Shalizi et al., 2006). Thus, transcription factor combinations such as GMT likely provide specificity to genome-wide epigenomic remodeling, but how they achieve tissue-specific transcriptional regulation remains unknown.

Gata4 is a zinc finger transcription factor capable of interacting with heterochromatin, but its ability to interact with regions containing DNA methylation is limited (Cirillo et al., 2002; Oda et al., 2013). Little is known regarding the ability of Tbx5 and Mef2c to bind to compact chromatin. One screen identified TBX5 as a factor capable of inducing DNA demethylation when expressed ectopically (Suzuki et al., 2017). While the function of Mef2c in this regard remains unclear, a study of the closely related factor Mef2d in photoreceptor cells found that it requires additional co-factors to access regions that do not encode strong consensus Mef-response motifs (Andzelm et al., 2015). Which of these factors, if any, functions to open closed chromatin in the context of direct reprogramming, and whether they require combinatorial interaction to do so, remains unknown.

Here, we investigated the genome-wide consequences of Gata4, Mef2c, and Tbx5 expression, alone and in combination, in cardiac fibroblasts as the cells underwent reprogramming towards a cardiomyocyte-like state. By combining single cell RNA-sequencing, ChIP-seq of GMT, and ATAC-seq analyses, we found that epigenomic and transcriptional changes occurred rapidly within the first 24-48 hours of reprogramming. Cells that adopted a trajectory toward the cardiac fate could largely be predicted by virtue of early gene expression changes and reprogramming factor expression. A machine learning model of gene expression changes as a function of transcription factor binding motifs in dynamic open chromatin regions identified new candidate factors involved in reprogramming. Although GMT are each capable of promoting chromatin
remodeling when expressed individually, we found that accessibility changes during direct reprogramming were primarily associated with Mef2c and Tbx5 binding only. Cooperative activity between Gata4, Tbx5, and Mef2c was evident as combinatorial expression resulted in refinement and facilitation of DNA binding by these factors compared to single factor expression, and combinatorial binding correlated with opening of chromatin particularly at cardiac loci.

RESULTS

Cardiac Reprogramming Occurs Rapidly and at Variable Rates

To determine the discrete temporal transcriptional response to reprogramming with GMT in the context of TGFβ and Wnt inhibition, we performed single cell RNA-sequencing during cardiac reprogramming of Thy1 positive (Thy1+) cells, largely representing fibroblasts, isolated from neonatal mouse hearts that encode an αMHC-GFP reporter activated during reprogramming (Ieda et al., 2010) (Figure 1A). We collected and analyzed 29,718 cells representing five time points after transduction with retroviruses encoding Gata4, Mef2c, and Tbx5 (days -1, 1, 2, 3, 7). We additionally collected cells sorted at day 14 using the aforementioned reporter.

Transcript information from all samples was aggregated and 14 distinct transcriptional signatures were identified (Figure 1B, Table S1) (Becht et al., 2018; Butler et al., 2018). Excluding the clusters that exclusively represent day -1 (clusters 6 and 11), all clusters included cells collected at each time point, highlighting the limited technical variability between our timepoints and the heterogeneous response to GMT (Figure 1C, D).

To understand the biological significance of the 6 main groups of cells identified through hierarchical clustering of our populations, we identified representative gene signatures for each cluster (Figure 1E). 7,395 genes were differentially expressed (p < 0.01, average log fold change > 0.3) across the time course (Table S1). Three populations represented non-fibroblast cell types present in the starting population: epicardium-derived (clusters 11 and 12, Lrrn4 and Mgp), endocardium (cluster 13, Emcn and Egfl7), and macrophages (clusters 8 and 9, Lyz2 and Spp1) (Figure 1D-E,
S1B) (Cavallero et al., 2015; Xiao et al., 2018). Notably, the expression of genes such as Lrrn4 and Emcn exclusively within endocardium and epicardium cells throughout our time course contradicts a previous report that suggested these genes are expressed at early stages in reprogramming cells but are subsequently repressed by GMT expression (Figure 1C-D) (Liu et al., 2017). Instead, our analysis suggests these distinct cell types persist in the population in low numbers and likely were not detected in the previous study due to a limitation in the number of cells captured.

Four signatures putatively represented various stages or outcomes of cardiac reprogramming. The initial stages (early iCMs) were identified by activation of genes such as Nid2 and Tnnt2 and incomplete repression of fibroblast-associated genes such as Lmcd1 and Ptgs2 (clusters 0 and 5; Figure 1E, S1B). This signature was present within 48 hours of GMT transduction, in agreement with previous reports (Liu et al., 2017; Sauls et al., 2018). The late stages (late iCMs) were marked by expression of cardiomyocyte-related genes such as Tnni3 and Myl7, and downregulation of the fibroblast genes such as Ptgs2 (cluster 2; Figure 1E). Unexpectedly, this cluster contained cells collected from days 3 (5% of day 3 cells) and 7 (26% of day 7 cells), as well as reporter-positive cells collected on day 14 (“+14r”; Figure 1C-D), suggesting a reprogrammed state can be acquired rapidly.

The two additional signatures appear to contain alternative reprogramming outcomes. The first exhibits expression of various genes associated with the cell cycle such as Cdk1 and Ccnb1 (clusters 4 and 10; Figure 1E). The second activates genes that become more robustly expressed in cluster 7, such as Mmp3 and Figf (Figure 1E, S1B). Cluster 7 is similar to clusters 11 and 12, which are found in the starting population, suggesting cluster 7 represents cells that do not acquire a cardiac fate nor enter the cell cycle (Figure 1D-E). Gene ontology (GO) analysis of genes activated in cluster 7 (compared to cluster 11, n=488 genes, average log fold change > 0.3, p < 1e-10) revealed biological processes associated with vasculature and blood vessel development (p = 1.76e-13 and 5.32e-13, respectively). These cells uniquely activate genes associated with vascular developmental processes such as Epas1, Figf, and Sox9 (Figure 1E, Table S1) (Achen et al., 1998; Lincoln et al., 2007; Tian et al., 1997).
They also continue to express genes associated with the starting fibroblast state, such as \textit{Dcn} and \textit{Tbx20} (Table S1).

To better understand the associations between identified clusters and establish a transcriptional trajectory of the reprogramming process, we next ordered clusters in pseudotime using Monocle (Cao et al., 2019). Clusters containing fibroblasts (cluster 6) and non-fibroblast cell types identified in the starting population (clusters 8, 9, 11, 12, and 13) were eliminated as they do not reflect the outcome of reprogramming. This analysis revealed a tree structure with distinct branches indicating three possible outcomes that originate from \textit{Slc1a6} positive cells (Figure 1F): one characterized by progressive activation of \textit{Mmp3}, another by activation of cardiac genes such as \textit{Tnni3}, and the third by markers of cell cycle progression (e.g. \textit{Ccnb1}) (Figure 1G). The inhibitory effect imposed by proliferation during cardiac reprogramming is consistent with the prior observation that continued proliferation prevents fibroblast reprogramming to a pluripotent state (Xu et al., 2013). Collectively, these data suggest that a reprogramming trajectory can be acquired within 48 hours of GMT transduction but that reprogramming progresses at variable rates in individual cells.

**Reprogramming Trajectory is Entered Quickly and Driven by GMT Transduction**

To understand how ectopic GMT expression may dictate the observed transcriptional trajectories, we assayed expression of Gata4, Mef2c, or Tbx5 in each cell by generating 5’ single cell RNA-sequencing data for 2,593 cells collected on day 1 of reprogramming. This approach circumvented the limitation of our initial analysis where the individual ectopic retroviral plasmids could not be distinguished because they each encode the same 3’ polyadenylation sequence. After eliminating the myeloid lineage, we identified 12 clusters within this population, confirming the prompt rate in which cells alter their transcriptional landscape in our system (Figure 2A). A pseudotime analysis again identified three main branches in the main trajectory as well as a separate group of clusters (clusters 6, 7, 11, and 12) that were unlinked from the main trajectory (Figure 2B).

The three branches identified in the main trajectory represent gene signatures analogous to those presented in Figure 1 based on a differential expression analysis:
cells that are reprogramming (A, 33%), those that are likely proliferating (B, 22%) and fibroblast-like cells (C, 29%) (Figure 2C, Table S2). Evaluation of GMT expression revealed their collective expression in branch A (Figure 2D), which contains cells that have activated markers of a cardiomyocyte fate (e.g. Pdlim3 and Smpx; Figure 2C-E) and downregulated genes associated with a fibroblast identity (e.g. Postn and Tbx20; Figure 2C, Table S2). Gata4 is expressed in cardiac fibroblasts and is therefore detected throughout the population; however, it is increased 1.6-fold in cluster 1 (branch A) compared to cluster 2 (branch C). GMT is also expressed in cluster 10, which lies within branch B and expresses Smpx and Pdlim3 at increased levels, suggesting expression of GMT can initiate reprogramming even if the cells enter a proliferative state (Figure 2C-D). While this type of reprogramming may not produce more advanced iCMs, it suggests proliferation does not prevent the initial stages of reprogramming (Liu et al., 2017). In contrast to branches A and B, the fibroblast-like cells that populate branch C exhibited only baseline levels of all three factors (Figure 2D). This analysis indicates that branch C represents fibroblasts that do not express ectopic GMT, rather than representing a newly acquired state driven by transduction with one or two factors. Therefore, unlike observations made during direct neural reprogramming, we do not detect the emergence of an alternative cell type in our experiments (Treutlein et al., 2016).

GMT expression was also detected within the unlinked trajectory in clusters 6 and 11 (Figure 2D). These clusters contain epicardial cells expressing Ddx4, Lrrn4, and Msln (Figure 2C, S2A, Table S2). Cluster 6 additionally upregulated early markers of the iCM trajectory such as Pdlim3 and Smpx, suggesting that, although unlinked from the main trajectory, this cell type may be capable of acquiring a cardiomyocyte-like gene expression signature upon transduction with GMT (Figure 2C, Table S2). A cell cycle-related phenomenon was observed in the epicardial cells similar to what we observed in branch B, as epicardial cells in cluster 11 entered the cell cycle and activated expression of Pdlim3 and Smpx (Figure 2C, S2B, Table S2).

To identify variables that may dictate progress in the reprogramming trajectory, we next compared the gene expression profiles of clusters 1 and 4 as they represent early, yet distinct iCM reprogramming states. Examination of GMT expression levels
found a statistically significant difference in Gata4 (p-value = 9.58e-45) and Tbx5 (p-value = 2.08e-15) between clusters 1 and 4, but not Mef2c (Figure 2E, S2B). Cluster 1 exhibited stronger upregulation of early markers of reprogramming (e.g. Cd24a, Smpx, and Tnnt2) and downregulation of fibroblast-associated genes (e.g. Postn, Sdpr, and Tbx20) (Figure 2G, Table S2). Therefore, while robust expression of Mef2c is required, this variation suggests lower levels of Gata4 and/or Tbx5 may limit the rate of reprogramming but nonetheless allow initiation of the process. While cluster 8 is most similar to clusters 1 and 4, it has not robustly activated markers of reprogramming but it has downregulated markers of the starting fibroblasts (Figure 2C, E-G). There was a significant difference in Mef2c expression between clusters 4 and 8 (p-value = 1.17e-08; Figure 2E), further supporting the necessity of robust expression of this gene.

**Chromatin Remodeling Occurs within 72 Hours of GMT Expression**

To identify the dynamics in chromatin accessibility underlying the aforementioned transcriptional changes, we performed ATAC-seq on αMHC-GFP* cells collected at five time points during reprogramming (days 2, 3, 7, 14, and 21), and compared regions of accessible chromatin to those detected in the starting fibroblast population. This analysis identified 100,691 total dynamic regions, which included a rapid gain of accessibility by day 2 of reprogramming at the early reprogramming marker gene Slc6a6 and cardiac Tnnt2 loci (Figure 3A, S3A-B, Table S3). Principal component analysis of the genome-wide chromatin accessibility data showed extensive chromatin remodeling by day 2, in agreement with the transcriptional dynamics presented in Figure 1 (Figure S3C).

To uncover factors that direct the most robust changes in chromatin accessibility, we performed hierarchical clustering on the 10,000 most differentially accessible regions identified during our time course and found eight primary patterns (Figure 3B). Approximately half of the most dynamic regions (n=4,480) identified in our time course lost accessibility during transdifferentiation, while the other half gained accessibility (n=5,520). Regardless of chromatin remodeling dynamics, the vast majority of changes occurred distal from transcriptional start sites, with dynamic regions underrepresented in promoter proximal regions (p = 2.2e-16; Figure S3D). The majority of regions that lost
accessibility exhibited this change within 3 days of GMT induction (clusters A1-4; Figure 3B, C). Motif enrichment analysis identified the TEAD family as most associated with loss of chromatin accessibility (Figure 3D, Table S3), specifically motifs for TEA transcription factors Tead1 and Tead4, which are both expressed throughout reprogramming (Table S4, Figure S3E).

In contrast to the similar dynamics observed in clusters that exhibited the strongest loss of accessibility, there were multiple distinct patterns associated with gain of accessibility. Cluster A5 demonstrated a gain in chromatin accessibility by day 2, but then exhibited a return towards the fibroblast accessibility state at later time points, suggesting that accessible chromatin at those sites was not stabilized (n=385; Figure 3B, C). This cluster showed limited enrichment of transcription factor sequence motifs, which may have prevented stable GMT binding similar to findings reported for Mef2c, Gata4, and FOXA2 where a lack of motif leads to transient sampling rather than stable binding (Figure 3D, Table S3) (Andzelm et al., 2015; Donaghey et al., 2018). Cluster A6 demonstrated an initial trend similar to cluster A5; however, the extent of accessibility loss at later time points was reduced (n=1,352; Figure 3B-C). Clusters A7 and A8 represent the majority of regions associated with a gain in accessibility, and the maximum gain in these regions was observed after day 3 (n=2,471 and n=1,212, respectively; Figure 3B-C). Regions in these clusters maintained higher levels of accessibility over the time course, compared to clusters A5 and A6, and also contained significant enrichment of multiple motif families (Figure 3D, Table S3). Thus, regions that transition from closed to open during cardiac reprogramming have a number of distinct patterns over time, including several that are only transiently open, and each is associated with different sequence motifs.

To assess the potential functional roles of each cluster, we annotated regions exhibiting changes in chromatin accessibility using GREAT (Figure 3E) (McLean et al., 2010). Regions that exhibited loss of accessibility were associated with the inflammatory response (cluster A2) and monocytes (cluster A3), supporting a previous report that demonstrates reprogramming is promoted by repression of inflammatory signaling pathways (Zhou et al., 2017). Clusters A7 and A8, which gain and maintain accessibility, were associated with cardiovascular terms such as cardiac and striated...
muscle development. Regions that did not maintain accessibility in cluster A5 were also associated with cardiac function (Figure 3E). It remains possible that while Tbx5 may transiently sample those sites during reprogramming, it requires developmentally regulated binding partners such as Eomes, or others, to initiate and/or stabilize the interactions with DNA that are not robustly detected in our system (McLane et al., 2013).

**Computational Modeling Reveals Additional Factors Involved in Cardiac Reprogramming**

In an effort to discern additional transcription factors that direct the initial stages of reprogramming, we devised a multivariate machine learning approach to predict which transcription factor sequence motifs are most associated with transcriptional changes that occur during the first 2 days of reprogramming (Figure 4A). We found a stronger correlation between sequence motif content of dynamic chromatin regions 2-500kb from the TSS (Pearson correlation = 0.37 between observed versus predicted fold-change, p-value < 0.05, t-test for correlation coefficient) as compared to within 2kb of the TSS (Pearson correlation = 0.22, p-value < 0.05, t-test for correlation coefficient), supporting a previous report that suggested chromatin dynamics proximal to the TSS were poor predictors of gene expression dynamics (Pliner et al., 2018). This model predicted 48 motifs significantly associated with transcriptional changes that occur during the fibroblast to day 2 time frame (Figure 4B). A lower correlation was detected when incorporating only the motifs of GMT that are within accessible chromatin at day 2 (correlation = 0.19), suggesting additional transcription factors are indeed involved in early transcriptional dynamics associated with reprogramming.

We ranked the identified motifs based on a net importance score (Figure 4B). A positive net importance score suggests an increase in transcription while a negative score suggests a repressive influence; a score close to zero indicates a mixed influence. This analysis found that the Tbx5 motif (T-box) was most associated with gene expression changes (Figure 4B). Notably, while Mef2 motifs (MADS) were also among the top ranked set of putative early regulators, Gata motifs were absent from this list, suggesting that Tbx5 and Mef2c are more influential than Gata4 in regulating gene
expression changes during the early stages of reprogramming. The lack of Gata4 motif detection in this prediction combined with ATAC-sequencing results suggests that its link to gene expression changes is weak. Motifs for Smads2/3/4 were also predicted to influence gene expression, supported by previous work that showed TGFβ inhibition positively influences reprogramming outcome (Ifkovits et al., 2014; Mohamed et al., 2017).

To reveal if the identified transcription factors target similar or distinct gene sets, we next performed hierarchical clustering to discover groups of motifs with similar relationships to gene expression changes (Figure 4C). Motifs clustered into two groups, one of which contained many motifs associated with cardiomyocyte development such as Mef2 and Tbx family motifs, as well as Sox, Fox, and SMAD motifs (Figure 4C, lower). The Tbx5 motif was most closely linked to changes at regions that also encode Tgif1 motifs, a TGFβ-induced transcriptional homeodomain-containing repressor (Figure 4C, bold) (Wotton et al., 1999). However, dynamics associated with the Mef2 motifs were not closely linked to any other motifs, suggesting Mef2c may function independently in the reprogramming context (Figure 4C, bold). Next we used a linear model to determine motif pairs whose influence on gene expression could not be explained by either of the motifs alone. Ranking transcription factor motifs by the total number of predicted interactions revealed strong enrichment for the motifs Tcfcp2l1, Sox4, and Hif1a, providing evidence that their role in reprogramming involves cooperative interactions with additional factors (Figure 4D, Table S5). These results further support a model in which multiple transcription factors jointly regulate gene expression dynamics during reprogramming.

To leverage the predictions made by the TF interaction model, we next examined the effect of shRNA-induced knockdown of selected factors on reprogramming efficiency at day 2. Of the 18 genes tested, 5 exhibited a statistically significant reduction in reprogramming efficiency (Sp1, Foxo1, Tcfp2l1, Tgif1, and Foxo1) while 3 improved reprogramming (Hif1a, Prdm1, and Smad3) (Figure 4E). Future studies of the remaining candidates may reveal additional factors that can enhance or serve as barriers of cardiac reprogramming, and their mechanisms of action.
Mef2c and Tbx5 Binding Is Associated with Changes in Chromatin Accessibility

Given that GMT drive cardiac reprogramming, we next performed ChIP-seq at day 2 of reprogramming to assess Gata4, Mef2c, and Tbx5 occupancy to dissect direct versus indirect consequences of their binding to DNA when introduced in combination. Unlike the ATAC-seq presented in Figure 3, this experiment was performed in immortalized neonatal cardiac fibroblasts to obtain sufficient numbers of cells (Figure S4A); therefore, we created a new, matched ATAC-seq dataset for integration with the ChIP-seq data. This analysis identified 5,100, 6,904, or 5,307 peaks for Gata4, Mef2c, or Tbx5, respectively (Figure 5A-B, S4B). The majority of Gata4 and Mef2c peaks were located further than 2 kilobases (kb) from the nearest transcription start site (TSS), while 45% of Tbx5 peaks were within 2kb of a TSS (Figure S4C).

To reveal relationships between reprogramming factor binding and chromatin accessibility, we performed hierarchical clustering on the merged region set (n=14,138) bound by Gata4, Mef2c, and/or Tbx5 during reprogramming, including changes in chromatin accessibility, and identified eight primary patterns (Figure 5B, S4B). The regions in two groups (clusters B1 and B3, n=2,630) were generally bound by all three factors. However, these clusters exhibited disparate chromatin dynamics. Cluster B1 contained regions that were accessible in the starting population, and their accessibility increased slightly by day 2 (Figure 5B). In contrast, regions within cluster B3 were inaccessible in the starting fibroblasts, and their accessibility increased (6.33-fold mean increase by day 2) (Figure 5B).

Two clusters contain regions bound by Mef2c alone (clusters B2 and B4; n=1,968 and n=2,240, respectively), each displaying opposing trends in chromatin accessibility (Figure 5B). While regions in cluster B2 lost accessibility on average, regions in cluster 4 experienced a 2.55-fold mean increase (Table S6). While the trend identified in cluster B4 suggests the machinery Mef2c requires to promote chromatin remodeling is active during reprogramming, we did not observe significant chromatin remodeling at regions bound by Tbx5 alone (cluster B7, n= 2,370) nor Gata4 alone (cluster B5, n=1,419) (Figure 5B). Cluster B7 exhibited little change in accessibility during reprogramming (0.64-fold mean change), while regions bound by Gata4 and Tbx5 together (cluster B6; n=2,129), exhibited a 3.84-fold mean increase in accessibility,
suggesting synergistic binding of Gata4 and Tbx5 has a positive impact on chromatin remodeling at those regions (Figure 5B).

We next identified potential non-GMT cofactors within these regions by searching for known motifs enriched within ChIP-seq peaks and summarized them based on TF family (Figure 5C, Table S5). As expected, top-ranked motifs correspond to families of reprogramming transcription factors; however, additional motif families were also significantly enriched, including those that bind bZip, Homeobox, and Forkhead proteins (Figure 5C, Table S5). These families include transcription factors such as Atf1/2/3/7, Fosl2, Jun, and Bach2 (bZip); Tgif1/2 and Meis1 (H-box); and Foxm1 (Forkhead), all of which are expressed during reprogramming (Figure S3E). This result suggests that combinatorial binding at dynamic chromatin is important beyond GMT.

Finally, we analyzed the binding of Gata4, Mef2c, and Tbx5 at regions that exhibited the most dynamic chromatin accessibility changes during reprogramming (regions from Figure 5B-C). We detected no enrichment of GMT binding at day 2 in regions that lose accessibility during reprogramming (clusters A1-4), while we detected statistically significant enrichment of GMT binding at day 2 in almost all regions that gained accessibility during reprogramming (clusters A6, A7, and A8; Figure 5D). Cluster A5 is the only ATAC-seq cluster that gained accessibility at day 2 of reprogramming without significant enrichment of binding by reprogramming factors, providing a potential explanation for why this cluster exhibits only a transient accessibility gain (Figure 5D, 3B-C). Taken together, these data suggest that chromatin accessibility dynamics directed by binding of GMT are context-specific.

Transcription Factor DNA Occupancy Defines Chromatin Accessibility Trends

To understand how GMT binding, individually and in combination, is related to changes in chromatin accessibility, we next performed ChIP-seq and ATAC-seq at day 2 on immortalized neonatal cardiac fibroblasts with single factors (SF) as well as pairs of factors ectopically expressed (double factor, DF) (Figure 6A). Overall, each individual factor’s binding pattern differed from that detected during reprogramming with all factors (AF) (Figure 6B). A large shift occurred for Gata4, whose binding became more similar to that of Tbx5 during AF reprogramming, supporting a previously reported
cooperative binding relationship between these two transcription factors in developing mouse and human cardiomyocytes (Ang et al., 2016; Luna-Zurita et al., 2016; Maitra et al., 2009). Mef2c exhibited a decidedly distinct binding pattern compared to Tbx5 and Gata4, but its binding was altered by the addition of Gata4 and Tbx5 (Figure 6B).

Hierarchical clustering of the merged region set when AF or an SF were detected, together with changes in chromatin accessibility, resulted in 8 distinct clusters (Figure 6C). Most clusters were driven by binding of a single factor (Figure 6C). Clusters C2 and C4-C6 represent clusters bound in single factor conditions that are refined by the addition of all factors (Figure 6C, S5A). Binding of a single factor in C2 and C5 was associated with a concomitant increase in chromatin accessibility. The single binding events refined by the addition of the other reprogramming factors coincide with overrepresentation of the single factor’s sequence motif (Figure S5B, Table S6). Regions in C5 exhibited an increase in chromatin accessibility only when Gata4 was present alone, which may suggest these regions act as regulatory elements in cell types for which Gata4 is involved but Tbx5 and Mef2c are not, such as at the TSS of the endothelial gene Lecam2 (Figure S5A). Similarly, Mef2c was sufficient to induce an increase in chromatin accessibility in a subset of regions within C2 that was abrogated by the addition of Gata4 and Tbx5 (mean fold changes, M=1.75, MG=1.24, MT=1.12, AF=0.84) (Figure 6C, Table S7).

Mef2c and Tbx5 SF binding events were also associated with a loss of chromatin accessibility (C4 and C6), suggesting their ability to interact with DNA and alter chromatin accessibility is context-dependent. C8 confirms the divergent response to Tbx5 binding as a subset of these regions were also bound by Gata4, and were associated with minimal changes in chromatin accessibility (Figure 6C). We did not identify a cluster in which Gata4 was independently capable of stably increasing chromatin accessibility, suggesting its function in the reprogramming context is downstream of epigenomic remodeling.

We noted unique trends in clusters C3 and C7. They are dominated by regions that exhibit binding of all three factors in the AF reprogramming condition, but limited binding in the SF conditions (Figure 6C). While Mef2c binding was sufficient to promote a 3.09-fold increase in chromatin accessibility in cluster C3, Tbx5 comparably directed a
3.08-fold increase in cluster C7, representing the greatest average increases in chromatin accessibility among regions included in this analysis (Figure 6C, Table S7). This suggests that chromatin changes induced by Mef2c and Tbx5 create a chromatin landscape amenable to the binding of the additional reprogramming factors, adding another layer of regulatory complexity.

We next ascertained the relationship between these clusters and the transcriptional signatures identified by single cell RNA-sequencing (Figure 1). To that end, we defined genes that represent early iCMs, late iCMs and untransduced fibroblasts (p < 0.0001; Table S1) and calculated the distance from the TSS to the closest dynamic region. Regions whose chromatin dynamics were largely associated with Mef2c binding (clusters C2 and C3) were significantly associated with genes marking “Early iCM” populations (p < 0.001; Figure 6D, left). Clusters C3 and C7 were associated with “Late iCM” genes and an increase in occupancy by all three reprogramming factors during reprogramming (p < 1e-10; Figure 6D, middle). This association between C3 and C7 with gene expression suggests that the “Late iCM” trajectory results from cooperative binding by all three reprogramming factors while Mef2c is associated with the initial gene expression changes that define early iCMs.

To identify chromatin dynamics associated with untransduced cells that do not reprogram, we next evaluated the distance between observed chromatin dynamics and genes that represent this trajectory. Indeed, we found that regions in cluster C5 that were bound by Gata4 only in the Gata4 SF condition were significantly closer to genes that represent the untransduced fibroblasts (p < 0.001; Figure 6D, right). The lack of chromatin accessibility changes in the DF conditions in C5 indicate that neither Mef2c (MG) nor Tbx5 (GT) bind to or prevent Gata4 from binding to or altering chromatin accessibility at these regions, supporting our conclusion that these cells represent an untransduced population that expresses and is regulated by endogenous levels of Gata4 (Figure 6A, C). Cumulatively, these data reveal the complexity of the mechanisms through which transcription factors influence each other, both enabling and refining one another’s ability to bind DNA and affect accessibility changes.

**Individual Factors Activate Transcription of Reprogramming Genes**
To understand the extent to which GMT synergy leads to gene expression changes during reprogramming, we next identified genomic regions bound by GMT that are proximal to differentially expressed genes. Ldb3 and Ptrf represent two genes whose expression increases in early iCMs by day 1 compared to the starting population and are bound by reprogramming factors; Ptrf is bound by Mef2c and Tbx5 in the AF context, while Ldb3 is bound by all three reprogramming factors in the AF context (Figure 7A, B). We predicted putative enhancer elements that may be responsive to GMT guided by the ChIP-seq and ATAC-seq data and designed reporter constructs with these regions to identify which factor(s) are sufficient to induce gene expression. Neonatal cardiac fibroblasts were concurrently transduced with these reporters as well as GMT, and reporter expression was assessed using FACS.

Despite evidence of binding by multiple factors at these sites during reprogramming, ectopic expression of a single factor was sufficient to induce reporter gene expression from both of these constructs. Mef2c, but neither Gata4 nor Tbx5, was sufficient to induce robust Ldb3-driven reporter expression, consistent with Mef2c’s binding to the endogenous Ldb3 locus independently, with subsequent binding of Tbx5 and Gata4 (Figure 7B, C). Conversely, while both Mef2c and Tbx5 were independently capable of binding to the endogenous Ptrf locus, only Tbx5 was sufficient to activate expression of the Ptrf-driven reporter when introduced alone (Figure 7B, D). While Mef2c had no discernable effect on reporter expression, the addition of Gata4 limited the ability of Tbx5 to induce reporter expression resulting in a mean decrease of 61% (uncorrected p-value = 0.006), demonstrating an example of the coregulatory refinement suggested by ChIP- and ATAC-seq data (Figure 5). Although both the Ldb3 and Ptrf genes are expressed in early iCMs, our results demonstrate that the role of each reprogramming factor is not limited to synergistic activation, but rather differs between these two loci, indicating context-specific effects in regulatory element usage for each transcription factor.

DISCUSSION

Here, we interrogated the transcriptional and epigenomic dynamics underlying direct cardiac reprogramming in an in vitro mouse cardiac fibroblast system, revealing
numerous insights into the mechanisms associated with the cell fate transition from a fibroblast toward a cardiomyocyte. Epigenomic and transcriptional changes occurred broadly within the first 72 hours, and cells destined to reprogram could largely be predicted by virtue of early gene expression dynamics and reprogramming factor expression. Single cell assays addressed long-standing questions regarding heterogeneity and response to combinations of reprogramming factors, clarifying existing interpretation of bulk transcriptome data sets. A machine learning approach revealed clusters of co-located transcription factor motifs within dynamically changing chromatin regions associated with coordinate gene expression changes, pointing to additional factors that may promote or inhibit reprogramming. Integration of GMT DNA occupancy with genome-wide chromatin accessibility and single cell RNA-sequencing in the setting of individual or combinations of transcription factors revealed an interdependency of their binding patterns and suggests possible mechanisms through which they facilitate successful reprogramming.

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 induced using a combination of Ascl1, Brn2, and Mytl (Treutlein et al., 2016), an alternative fate characteristic of skeletal muscle was observed. In contrast, for cells that were successfully transduced with GMT, no major alternative fates, compared to starting cell types, were observed. Alternatives may be limited in the cardiac setting, because, as we show here, GMT binding is refined when they are expressed combinatorially, perhaps focusing binding events on cardiac loci. By contrast, Ascl1’s binding is not altered by the addition of other neural reprogramming factors such as Brn2 and Mytl1, suggesting its binding is unrestrained and may occur at regulatory elements employed in the development of multiple cell types (Wapinski et al., 2013).

While reports of direct reprogramming were first documented many years ago, the tools available to dissect the precise molecular mechanism of reprogramming were limited. Advances in single cell RNA-sequencing have created avenues to identify the path a single cell can take to its endpoint, and identify the molecular determinants of these trajectories (Cacchiarelli et al., 2018; Schiebinger et al., 2019). Simultaneous
advances in machine learning have improved the information gleaned from single cell RNA-sequencing data, as well as the ability to correlate changes between gene expression and chromatin remodeling (Cao et al., 2018; Deng et al., 2019; Eraslan et al., 2019; Lopez et al., 2018; Way and Greene, 2018; Welch et al., 2017). The observation that the vast majority of fibroblasts that expressed GMT proceeded into the induced cardiomyocyte trajectory suggests a higher efficiency among GMT-expressing cells than previously recognized and suggests that reprogramming efficiency might be improved by increasing the proportion of fibroblasts in which all three factors are ectopically expressed. Furthermore, our analysis suggests that other cell types, such as epicardial cells, have the potential to be partially reprogrammed, although they remain dissimilar to fibroblast-derived induced cardiomyocytes.

In conclusion, we have developed a comprehensive genomic assessment of transcription factor binding, chromatin state, and transcriptional changes, that reveals the molecular complexity involved in direct cardiac reprogramming. Mechanistic insights provided by integration of multiple datasets have started to reveal how lineage-enriched transcription factors can induce cell fate transitions in a combinatorial fashion, thereby achieving specificity of gene regulation.
Acknowledgments

We thank members of the Srivastava lab for helpful discussions; T.R. Roberts, G. Maki, and K. Claiborn for critical comments and editing on the manuscript; A.G. Williams, T.A. Friedrich, and S. Thomas in the Gladstone Bioinformatics Core; Y. Hao and J. McGuire in the Gladstone Genomics Core; M. Cavrois, M. Gesner, and N. Raman in the Gladstone Flow Cytometry Core; S. Elmes in the UCSF Flow Cytometry Core; E. Chow and D. Bogdanoff in the UCSF Center for Advanced Technology Core; and C. Benitez and I. Espineda in the Gladstone Laboratory Animal Resource Center. All protocols concerning animal use were approved by the IACUC at the University of California San Francisco and conducted in strict accordance with the NIH Guide for the Care and Use of Laboratory Animals. N.R.S was supported by a National Science Foundation Graduate Research Fellowship, R01 Research Supplement to Promote Diversity in Health-Related Research, and a Ruth L. Kirschstein NRSA Institutional Research Training Grant. C.A.G. is an HHMI fellow of the Damon Runyon Cancer Research Foundation (DRG-2206-14). K.S.P. was supported by NHLBI/NIH grants P01 HL089707 and UM1 HL098179. D.S. is supported by NHLBI/NIH grants R01 HL057181, P01 HL089707, and UM1 HL098179; the Roddenberry Foundation; the L.K. Whittier Foundation; and the Younger Family Fund. This work was also supported by NIH/NCRR grant C06 RR018928 to the Gladstone Institutes and NIH P30 AI027763, NIH S10 RR028962, and the James B. Pendleton Charitable Trust to the Gladstone FACS core.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar, 2002) and are accessible through GEO Series accession number GSE131328 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131328).

Author Contributions

Declaration of Interests
D.S. is a co-founder and member of the board of directors of Tenaya Therapeutics. D.S., K.I.,
K.S.P., and T.M.A.M. have equity in Tenaya Therapeutics.

FIGURE LEGENDS

Figure 1. Single Cell Expression Analysis of Direct Cardiac Reprogramming.

(A) Schematic of reprogramming system and main reprogramming milestones.

(B) UMAP visualization of the clustering of 29,718 cells into 14 clusters (0-13) identified by a graph-based clustering approach.

(C) UMAP visualization of cells in (B) colored by collection time (days).

(D) Stacked bar plot indicating the relative contribution of cells from each time point (as shown in C) to each cluster (as shown in B). Dendrogram is generated based on a distance matrix constructed in principal component analysis (PCA) space.

(E) Heatmap showing expression of top 20 differentially expressed genes for each cluster. Each row shows the fold change in expression of a single gene compared to the population mean of that gene. Red indicates higher expression; blue lower. Clusters (top bar, colors as in (B)) are ordered according to dendrogram in (D). Two representative marker genes per cluster are labelled at right. Putative cell identities (bottom bar) were determined based on differential expression.

(F) Pseudotime trajectory of cells from clusters 0-5, 7, and 10. Cell color is based on cluster color in (B).

(G) Expression [log10(UMI+0.1)] of branch marker genes (Tnni3, Mmp3, Ccnb1, and Slc1a6) visualized in pseudotime trajectory plots from (F).

See also Figure S1 and Table S1.
Figure 2. Cardiac Reprogramming Trajectory Is Entered Rapidly.

(A) UMAP visualization of 2,593 cells collected at day +1 and colored by clusters.

(B) Pseudotime trajectories of all cells in (A). Color indicates progression in pseudotime space. Grey indicates a disjointed trajectory.

(C) Dot plot showing expression of top marker genes from each of the 12 clusters in (A) based on specificity calculated by Moran’s I test. Color represents the gene’s average expression level [log10(UMI+0.1)] and size the percentage of cells expressing the gene. Background color indicates branches from (B).

(D, E) Expression [log10(UMI+0.1)] of (D) branch markers Ccna2, Pdlim3, and Tcf21 and of (E) Gata4, Mef2c, and Tbx5, overlayed on the trajectory plot in (B).

(F, G) Violin plots depicting normalized UMI levels for (F) Gata4, Mef2c, and Tbx5 and (G) Pdlim3, Tbx20, and Tnnt2 in clusters 1, 4, and 8, color-coded as in (A). Stars indicate negative binomial adjusted p-values.

See also Figure S2 and Table S2.

Figure 3. Cardiac Reprogramming Initiates Rapid and Distinct Patterns of Chromatin Accessibility Changes.

(A) Gain of ATAC-seq signal in iCMs (purple) harvested at indicated days of GMT-induced reprogramming compared to fibroblasts (black) near the early reprogramming marker gene Slc6a6 (left) and cardiomyocyte gene locus Tnnt2 (right).

(B) Hierarchical clustering of 10,000 regions with most differentially accessible chromatin status in αMHC-GFP+ iCMs harvested at indicated days of reprogramming compared to fibroblasts infected with dsRed control retrovirus. Red indicates accessibility gain, blue indicates decrease.
(C) Medoid plots showing ATAC-seq tag density over time at dynamic regions from each cluster in (B) representative of overall trends. Gray region represents the 95% confidence interval of the standard error of the mean signal profile of the representative region.

(D) Tables listing transcription factor families with motifs significantly enriched ($p \leq 1e^{-19}$) within dynamic regions from each cluster compared to all stably accessible (non-dynamic) regions. P-values listed are from the top ranked motif from each transcription factor family. Complete motif enrichment results are available in Table S3.

(E) Bar charts showing top three ranked biological process terms enriched in dynamic regions from each cluster compared to all stably accessible (non-dynamic) regions. Enriched annotations were determined using GREAT (McLean et al., 2010).

See also Figure S3, Table S3, and Table S4.

**Figure 4. Model of Transcription Factor Association with Changing Rate of Gene Expression over Time.**

(A) Schematic representation of input features for the multivariate machine-learning approach used to predict transcription factors influencing gene expression changes during reprogramming.

(B) Bar chart displaying net importance scores for 48 transcription factor motifs significantly associated with gene expression changes observed during the fibroblast-to-day-2 reprogramming transition. Motif family is listed in parentheses.

(C) Heatmap of clustered gene-association signatures for identified motifs. Rows represent transcription factor motifs; columns are genes associated with these motifs.
Red indicates enrichment of the motif nearby a gene, while blue indicates depletion of the motif; white indicates no predicted influence of the transcription factor on the gene.

(D) Total predicted interactions for each candidate transcription factor predicted to synergistically interact with at least one additional candidate. Black bars indicate candidates selected for knockdown experiments.

(E) Reprogramming outcomes (%αMHC-GFP+ cells) following knockdown of 18 candidate transcription factors at day 2 of reprogramming, based on machine-learning based predictions for cooperative interactions. Values indicated are the mean of three replicates. Paired t-test uncorrected p-value < 0.05 indicated by “*”.

See also Table S4 and Table S5.

Figure 5. Chromatin Accessibility Dynamics Associated with Gata4, Mef2c, and Tbx5 Occupancy during Reprogramming.

(A) ChIP-seq profiles display GMT binding at day 2 of reprogramming near Nid2 (left panel) and Actn2 (right panel). ChIP-seq track height is normalized to sample read depth of condition with highest number of reads. ATAC-seq track height is normalized to sample read depth of condition with highest number of mapped read pairs. Tracks are representative of read density from merged biological replicates (Gata4 ChIP, n=6; Mef2c ChIP, n=4; Tbx5 ChIP, n=4; for all ATAC-seq conditions n=3).

(B) Heatmap displays hierarchical clustering of ChIP-seq peaks. Grey scale displays average tag density across replicates, normalized to input, with black indicating an increase in tag density in sample over input. At right, co-clustered changes in accessibility. Blue-red color scale indicates tag density of ATAC-seq at day 2 of reprogramming, normalized to fibroblast. Red indicates a relative increase in accessibility, while blue indicates a decrease.

(C) Tables display transcription factor families with motifs significantly enriched within GMT-bound regions from each cluster, compared to stably accessible (non-dynamic)
regions. P-values listed are from the top ranked motif within each transcription factor family, calculated using the cumulative hypergeometric distributions. (Full enrichment results available in Table S6.)

(D) Bar plot displays the percentages of accessible chromatin regions (ATAC-seq peaks) from each cluster (A1-A8 in Figure 3B), bound by Gata4, Mef2c, or Tbx5. Statistical significance was determined using Fisher exact test for overrepresentation of overlapping ChIP-seq peaks in one cluster compared to all dynamic region clusters.

See also Figure S4 and Table S6.

Figure 6. Chromatin Accessibility Dynamics Associated with Gata4, Mef2c, and Tbx5 Occupancy Following Independent and Combinatorial Expression.

(A) Profiles display ChIP-seq signal for GMT in single factor (SF) and all factor (AF) conditions, and ATAC-seq signal in SF, double factor (DF), and AF conditions, near the early reprogramming marker gene locus, Ehd4. Grey rectangle highlights region bound by Gata4, Mef2c, and Tbx5 in AF condition, without binding by any of these factors in SF conditions. Blue rectangles highlight regions bound primarily by Mef2c when ectopically expressed alone, but bound by Gata4, Mef2c, and Tbx5 when all factors are expressed. Profiles represent read density from merged biological replicates normalized to read depth (Gata4 AF ChIP, n=6; Mef2c and Tbx5 AF ChIP, n=4; SF ChIP conditions, n=3; all ATAC-seq conditions, n=3).

(B) Principal component analysis of all ChIP-seq replicates based on 500 most variable regions among all conditions. Gata4 ChIP-seq samples are shown in orange, Mef2c in blue, Tbx5 in green, and input in black. Single factor (SF) expression conditions indicated by a triangle (∆) and all factor (AF) expression indicated by a circle (○).

(C) Heatmap displays hierarchical clustering of ChIP-seq peaks, with grey scale displaying average tag density across replicates, normalized to input (left). Black indicates an increase in tag density in sample over input. At the right, co-clustered tag
density of ATAC-seq, normalized to fibroblast. Red indicates a relative increase in accessibility, while blue indicates a decrease. Co-clustering reveals 8 distinct patterns (C1-C8). Conditions include exogenously expressed single factors (SF), double factors (DF), and all reprogramming factors (AF), with unsorted cells harvested at day 2 of reprogramming.

(D) Violin plots show distribution of distances from ChIP-seq peaks in each cluster to nearest TSS of differentially expressed genes in cells from “Early iCM”, “Late iCM”, and untransduced fibroblast single cell RNA-sequencing clusters. Internal lines indicate median values. Wilcoxon p-values were determined per cluster, by comparing distances to TSSs of DE genes with distances to all TSSs, and adjusted for multiple testing using Bonferroni correction.

See also Figure S5 and Table S7.

**Figure 7. Individual Reprogramming Factors Activate Transcription at GMT-bound Regulatory Regions of Early Reprogramming Genes**

(A) Violin plots depicting normalized UMI levels for early iCM marker genes Ldb3 and Ptrf in fibroblasts and iCMs at day 1 of reprogramming. Fold change (fc) listed above plot.

(B) Profiles display ChIP-seq and ATAC-seq signal near Ldb3 and Ptrf loci in the setting of single factor (SF), double factor (DF) or all factor (AF) conditions. Peak calls indicated above ChIP-seq profiles in gray. Blue rectangles highlight putative regulatory regions investigated by reporter assays in panels (C, D).

(C) Ldb3-mKate and (D) Ptrf-mKate reporter activation at day 1 of reprogramming with single, double, and all factor (AF) infections. Values displayed are means of four replicates. Error bars indicate standard deviation. T-test uncorrected p-value thresholds as indicated.
SUPPLEMENTAL FIGURE LEGENDS

Figure S1., Related to Figure 1.

(A) Representative fluorescence-activated cell sorting (FACS) plot for αMHC-GFP+ cells at week 2 of reprogramming.

(B) Expression [log10(UMI+0.1)] of marker genes for each cluster visualized in UMAP plots. Numbered clusters from Figure 1B listed at right for reference.

Figure S2., Related to Figure 2.

(A) Expression [log10(UMI+0.1)] of epicardial markers visualized in a UMAP trajectory plot.

(B) Violin plots representing normalized UMI levels of genes presented in Figures 2F-G for all clusters.

Figure S3., Related to Figure 3.

(A) Bar graph indicates total number of ATAC-seq peaks identified for each replicate (1-3). Sorted αMHC-GFP+ cells were assayed at each time point and in sorted dsRed-positive fibroblasts infected with a control dsRed retrovirus.

(B) Bar graph indicates total number of concordant ATAC-seq peaks for each condition and total number of replicate concordant peaks overlapping between conditions. Concordant peaks met an FDR threshold of < 0.05 in at least two out of three replicates.

(C) Principal component analysis of all ATAC-seq replicates (n=3) for each condition based on 10,000 most differentially accessible regions among all reprogramming time points compared with fibroblasts.

(D) Histogram shows distribution of regions within each dynamic cluster (A1-A8) and in a set of non-dynamic regions that are stably accessible in fibroblasts and in all reprogramming conditions, binned based on the distance from each peak to the nearest transcription start site (TSS). Dynamic regions were underrepresented within 2kb of a
TSS compared with non-dynamic regions accessible in all conditions (Fisher's p-value = 2.2e-16 for both up- and downstream bins).

(E) Heatmap displays expression levels for transcription factors with motifs with highest enrichment in dynamic accessibility clusters (A1-A8) compared with non-dynamic region set stably accessible in all conditions. Values indicate log2 of the mean normalized CPM across all replicates (n=3) from bulk RNA-seq completed on sorted αMHC-GFP+ cells collected at each time point during reprogramming.

**Figure S4., Related to Figure 5.**

(A) Representative fluorescence-activated cell sorting (FACS) plot for αMHC-GFP+ cells at day 2 of reprogramming, originating from immortalized neonatal cardiac fibroblasts.

(B) Heatmap displays tag densities of ChIP-seq sample and inputs, for all replicates. Counts were normalized for differences in sequencing depth between samples using upper quartile normalization separately for the ChIP and the input samples of each transcription factor. Columns represent independent replicates. Rows retain order shown in **Figure 5B**.

(C) Bar plot displays binned distribution of ChIP-seq peaks from nearest transcription start site (TSS). We identified a lower proportion of dynamic regions (all clusters) within 2kb up- and downstream of the nearest TSS compared with stably accessible, non-dynamic regions (Fisher odds ratio = 0.2 and 0.07, respectively; p-value = 2.2e-16 for both comparisons).

**Figure S5., Related to Figure 6.**

(A) Browser tracks display ChIP-seq in single factor and all factor conditions, and ATAC-seq signal in single factor, double factor, and all factor conditions, near leukocyte-endothelial cell adhesion molecule 2 (**Lecam2**). Track height was normalized to sample read depth within each assay. Orange rectangle highlights region from cluster 5 with Gata4 binding in the single factor condition which is ameliorated when all factors are present.
(B) Bar graph indicates percentage of regions within each cluster (C1-C8) that contain Gata4 (orange), Mef2c (blue), or Tbx5 (green) motifs. P-values were calculated using the cumulative hypergeometric distributions of motif occurrence in clustered region set compared with motif occurrence in non-dynamic regions stably accessible in fibroblasts and reprogramming conditions. Full motif enrichment results available in Table S7.

(C) Bar chart displays mean accessibility fold changes based on normalized tag densities in single infection (SF) conditions and in all factor (AF) condition, compared with fibroblasts, for regions within each cluster.
SUPPLEMENTAL TABLE LEGENDS

Table S1. Differential Gene Expression for Each Cluster, Related to Figure 1.
Contains average log fold changes and p-values for all differentially expressed genes from negative binomial tests used to determine identities of populations in clusters in Figure 1B, and designates top 20 marker genes per cluster used to generate Figure 1E.

Table S2. Differential Gene Expression for Each Cluster, Related to Figure 2.
Contains average log fold changes and p-values for all differentially expressed genes from negative binomial tests used to determine identities of populations in clusters in Figure 2A, and results of Moran’s I tests, including top two marker genes per cluster used in Figure 2C. Also contains average fold changes and p-values for all differentially expressed genes from pairwise tests of cluster 2 vs cluster 1, cluster 1 vs cluster 4, and cluster 4 vs cluster 8.

Table S3. Dynamic ATAC-seq Regions, Related to Figure 3.
Table containing 100,691 regions dynamic between fibroblasts and αMHC-GFP+ cells collected at reprogramming time points day 2, day 3, week 1, week 2, and week 3. Second sheet contains clustered region identities from Figure 3B. Third sheet contains known motif enrichment results comparing regions in each cluster to stably accessible background region set, summarized in Figure 3D.

Table S4. Bulk RNA-seq Gene Expression during Reprogramming, Related to Figure 3 and Figure 4.
Table includes normalized CPM values for each replicate over time course. Includes differential expression results between conditions.

Table S5. Predicted Co-regulatory Interactions of Transcription Factor Candidates, Related to Figure 4.
Table includes total number of cofactors predicted by model for candidate regulators of reprogramming, with totals listed separately for predicted interactors with positive and negative net importance scores.
Table S6, Related to Figure 5. Table contains known motifs enriched in regions clustered by Gata4, Mef2c, Tbx5 occupancy and chromatin accessibility at day 2 of reprogramming, summarized in Figure 5C.

Table S7, Related to Figure 6. Known motifs enriched in regions clustered by Gata4, Mef2c, Tbx5 occupancy and chromatin accessibility during combinatorial exogenous expression of individual factors. Fold change in ChIP-seq signal over input and ATAC-seq signal over fibroblast control for all regions depicted in Figure 6C.

REFERENCES


Figure 1.
Stone, Gifford, et al.

A

B

C

D

E

F

G

Retroviruses
Gate4

Chemical Inhibitors
TGFβ

First

Gating

Day

−1

+1

+2

+3

+7

+14

14 Clusters
29,718 Cells

myeloid

late iCM

epi/endo

early iCM

alternatives

fibroblasts

Pseudotime Trajectory

Present at Start

Alternative Reprogramming Outcome

Early iCM

Late iCM
Figure 2.
Stoner, Gifford, et al.

A  
Day +1
5' Single Cell RNA Sequencing
2,593 Cells

B  
Pseudotime Trajectory

C  
Gene   
<table>
<thead>
<tr>
<th>epi</th>
<th>B</th>
<th>C</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bax1</td>
<td>Tbx5</td>
<td>Smnx</td>
<td>Pdlim3</td>
</tr>
<tr>
<td>Top2a</td>
<td>Ccn3a2</td>
<td>Cenpx</td>
<td>Cenpx</td>
</tr>
<tr>
<td>Cenpf</td>
<td>Ckap2l</td>
<td>Plk1</td>
<td>Cenpe</td>
</tr>
<tr>
<td>Ccna2</td>
<td>Top2a</td>
<td>Pdlim3</td>
<td>Smpx</td>
</tr>
<tr>
<td>Top2a</td>
<td>Cenpx</td>
<td>Cenpe</td>
<td>Ccna2</td>
</tr>
</tbody>
</table>

D  
Gata4  Mef2c  Tbx5

E  
Gata4  
| 1 | 4 | 8 | ns |

F  
Pdlim3  
| 1 | 4 | 8 | * |

G  
Cd24a  Postn  Sdpr  Smnx  Tbx20  Tnnt2
Figure 3
Stone, Gifford, et al.

A

Fibroblast
Week 1
Week 2
Week 3
Day 2
Day 3

10,000 Most Differentially Accessible Regions

Cluster
A1 n=1,125
A2 n=1,318
A3 n=1,556
A4; n=481
A5; n=385
A6 n=1,352
A7 n=2,471
A8 n=1,212

B

Cluster

Accessibility Trend

Motif Family p-value
TEA 1e-49
Zf 1e-46
NR 1e-42
H-box 1e-40
bHLH 1e-36
TEA 1e-52
bZIP 1e-45
H-box 1e-43
CTF 1e-40
MAD 1e-40
TEA 1e-65
H-box 1e-62
HMG 1e-53
MAD 1e-52
bZIP 1e-51
TEA 1e-25
MAD 1e-21
Runt 1e-21
T-box 1e-19
MADS 1e-48
H-box 1e-43
bHLH 1e-43
NR 1e-38
MADS 1e-178
T-box 1e-121
H-box 1e-111
Zf 1e-105
bHLH 1e-99
MADS 1e-79
bHLH 1e-71
Zf 1e-57
T-box 1e-51
NR 1e-46

C

Log2(tag density)

D

E

Nearby Gene Ontology

Protein Localization
Metabolic Process
Biosynthetic Process

Peptide Cross-Linking
Inflammatory Response
Positive Regulation of Ras

Mammary Duct Growth
Positive Regulation of Ras
Metabolic Process

Leukotriene Production
Monocyte Chemotaxis
Smooth Muscle Adhesion

Cardiac Muscle Proliferation
Cardiac Ventricle Formation
Cardiac Chamber Formation

No significant terms.

Cardiac Muscle Development
p38MAPK Cascade
Actomyosin Organization

Visceral Muscle Development
Striated Muscle Development
Actin-Myosin Filament Sliding
Figure 4.
Stone, Gifford, et al.

A Transcription Factor Motifs

Accessibility

Transcriptional Changes

Predict Important Transcription Factors

B Individual Importance Ranking

<table>
<thead>
<tr>
<th>Motif</th>
<th>Net Importance Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tbx5 (Tbox)</td>
<td>0.5</td>
</tr>
<tr>
<td>Smad4 (MAD)</td>
<td>0.4</td>
</tr>
<tr>
<td>Smad2 (MAD)</td>
<td>0.3</td>
</tr>
<tr>
<td>Foxm1 (Forkhead)</td>
<td>0.2</td>
</tr>
<tr>
<td>Met2 (MADS)</td>
<td>0.1</td>
</tr>
<tr>
<td>Elf1 (ETS)</td>
<td>0</td>
</tr>
<tr>
<td>Tbx20 (Tbox)</td>
<td>−0.1</td>
</tr>
<tr>
<td>Six1 (Homeobox)</td>
<td>−0.2</td>
</tr>
<tr>
<td>Mel2 (MADS)</td>
<td>−0.3</td>
</tr>
<tr>
<td>Sp1 (Zf)</td>
<td>−0.4</td>
</tr>
<tr>
<td>Klf4 (Zf)</td>
<td>−0.5</td>
</tr>
<tr>
<td>E2f4 (E2F)</td>
<td>−0.6</td>
</tr>
<tr>
<td>E2f6 (E2F)</td>
<td>−0.7</td>
</tr>
<tr>
<td>E2f1 (E2F)</td>
<td>−0.8</td>
</tr>
<tr>
<td>E2f3 (E2F)</td>
<td>−0.9</td>
</tr>
<tr>
<td>E2f2 (E2F)</td>
<td>−1</td>
</tr>
<tr>
<td>Tcf4 (HMG)</td>
<td>0.1</td>
</tr>
<tr>
<td>Zfx (Zf)</td>
<td>0.2</td>
</tr>
<tr>
<td>Ebf1 (EBF)</td>
<td>0.3</td>
</tr>
<tr>
<td>Zfp809 (Zf)</td>
<td>0.4</td>
</tr>
<tr>
<td>Prdm1 (Zf)</td>
<td>0.5</td>
</tr>
<tr>
<td>Rfx1 (HTH)</td>
<td>0.6</td>
</tr>
<tr>
<td>Rfx5 (HTH)</td>
<td>0.7</td>
</tr>
<tr>
<td>Stat1 (Stat)</td>
<td>0.8</td>
</tr>
<tr>
<td>Ctcf (Zf)</td>
<td>0.9</td>
</tr>
<tr>
<td>Yy1 (Zf)</td>
<td>1</td>
</tr>
<tr>
<td>Bach1 (bZIP)</td>
<td>−0.1</td>
</tr>
<tr>
<td>Mafk (bZIP)</td>
<td>−0.2</td>
</tr>
<tr>
<td>Gilz (Zf)</td>
<td>−0.3</td>
</tr>
<tr>
<td>Maz (Zf)</td>
<td>−0.4</td>
</tr>
<tr>
<td>Sp1 (Zf)</td>
<td>−0.5</td>
</tr>
<tr>
<td>Sox15 (HMG)</td>
<td>−0.6</td>
</tr>
<tr>
<td>Sox6 (HMG)</td>
<td>−0.7</td>
</tr>
<tr>
<td>Sox4 (HMG)</td>
<td>−0.8</td>
</tr>
<tr>
<td>Sox3 (HMG)</td>
<td>−0.9</td>
</tr>
<tr>
<td>Sox2 (HMG)</td>
<td>−1</td>
</tr>
<tr>
<td>Bach2 (bZIP)</td>
<td>0.1</td>
</tr>
<tr>
<td>Bach1 (bZIP)</td>
<td>0.2</td>
</tr>
<tr>
<td>Klf9 (Zf)</td>
<td>0.3</td>
</tr>
<tr>
<td>Smad3 (MAD)</td>
<td>0.4</td>
</tr>
<tr>
<td>Smad2 (MAD)</td>
<td>0.5</td>
</tr>
<tr>
<td>Bach2 (bZIP)</td>
<td>0.6</td>
</tr>
<tr>
<td>Bach1 (bZIP)</td>
<td>0.7</td>
</tr>
<tr>
<td>Klf4 (Zf)</td>
<td>0.8</td>
</tr>
<tr>
<td>Klf9 (Zf)</td>
<td>0.9</td>
</tr>
<tr>
<td>Smad3 (MAD)</td>
<td>1</td>
</tr>
</tbody>
</table>

C Clustered Gene Associations

D Predicted Synergistic Interactions

E %MHC-GFP Scrambled

<table>
<thead>
<tr>
<th>Putative Interactor</th>
<th>Predicted Synergistic Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tcf4</td>
<td>10</td>
</tr>
<tr>
<td>Foxo1</td>
<td>9</td>
</tr>
<tr>
<td>Foxm1</td>
<td>8</td>
</tr>
<tr>
<td>Foxa2</td>
<td>7</td>
</tr>
<tr>
<td>Foxo1</td>
<td>6</td>
</tr>
<tr>
<td>Foxm1</td>
<td>5</td>
</tr>
<tr>
<td>Foxa2</td>
<td>4</td>
</tr>
<tr>
<td>Foxo1</td>
<td>3</td>
</tr>
<tr>
<td>Foxm1</td>
<td>2</td>
</tr>
<tr>
<td>Foxa2</td>
<td>1</td>
</tr>
<tr>
<td>Foxo1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Putative Interactor</th>
<th>%MHC-GFP Scrambled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp1</td>
<td>100</td>
</tr>
<tr>
<td>Foxo1</td>
<td>90</td>
</tr>
<tr>
<td>Foxm1</td>
<td>80</td>
</tr>
<tr>
<td>Foxa2</td>
<td>70</td>
</tr>
<tr>
<td>Foxo1</td>
<td>60</td>
</tr>
<tr>
<td>Foxm1</td>
<td>50</td>
</tr>
<tr>
<td>Foxa2</td>
<td>40</td>
</tr>
<tr>
<td>Foxo1</td>
<td>30</td>
</tr>
<tr>
<td>Foxm1</td>
<td>20</td>
</tr>
<tr>
<td>Foxa2</td>
<td>10</td>
</tr>
<tr>
<td>Foxo1</td>
<td>0</td>
</tr>
</tbody>
</table>

* p < 0.05
Figure 5.
Stone, Gifford, et al.

A

ChIP-seq

Gata4
Mef2c
Tbx5
Input

ATAC-seq

Fibroblast
Day 2

CpG Island
Nid2
Actn2

20 kb

B

Cluster
Gata4
Mef2c
Tbx5
Accessibility Change

B1
n=671

B2
n=1,968

B3
n=1,959

B4
n=2,240

B5
n=1,419

B6
n=2,129

B7
n=2,370

B8
n=1,382

14,138 GMT-bound Regions

C

Motif Family | p-value
---|---
MADS | 1e-100
Zf | 1e-26
bZIP | 1e-26

MADS | 1e-211
bZIP | 1e-150

MADS | 1e-303
T-box | 1e-112
H-box | 1e-40
Zf | 1e-38

MADS | 1e-371
bZip | 1e-52

Zf | 1e-211

T-box | 1e-271
Zf | 1e-152
H-box | 1e-90
Forkhead | 1e-19

T-box | 1e-248
Zf | 1e-40

Zf | 1e-325
H-box | 1e-20
T-box | 1e-16

D

% ATAC-seq Peaks Intersecting GMT Binding Site

0 5 10 15 20 25 30

0 5 10 15 20 25 30

ATAC-seq Cluster

Gata4
Mef2c
Tbx5

*p<0.0001
Figure 6.
Stone, Gifford, et al.

A

20 kb

Cluster

28,424 GMT-bound Regions

B

Gata4
Mef2c
Tbx5
Input

AF
SF

PC2: 10% Variance

PC1: 79% Variance

C

ChIP-seq
ATAC-seq

Cluster

C1
n=1,808

C2
n=3,501

C3
n=2,929

C4
n=2,374

C5
n=6,736

C6
n=3,203

C7
n=4,348

C8
n=3,524

D

Early iCM Genes
(n=23)

Late iCM Genes
(n=157)

Untransduced Fibroblast Genes
(n=35)
Figure 7.
Stone, Gifford, et al.

A
B

C
D

Ldb3 Reporter Activation

Ptrf Reporter Activation

%Ldb3-mKate+

%Ptrf-mKate+

0 10 20 30 40 50

0 10 20 30 40 50

G M T GM GT MT GMT

Single Factors Double Factors AF

G M T GM GT MT GMT

Single Factors Double Factors AF

p < 0.002
p < 0.05
n=4

p < 0.001
p < 0.01
n=4
SUPPLEMENTAL INFORMATION

Supplemental information includes 5 figures and 7 tables.
Figure S1, Related to Figure 1.
Stone, Gifford, et al.

A

B

Cluster Reference

GFP+
6.36%
Figure S2, Related to Figure 2.
Stone, Gifford, et al.

A

B

Expression

Cluster
Figure S3, Related to Figure 3.
Stone, Gifford, et al.

A)

B)

C)

D)

E)
Figure S4, Related to Figure 5.
Stone, Gifford, et al.

A

B

C

Cluster

Gata4 n=6
Mef2c n=4
Tbx5 n=4
Input n=6

B1
B2
B3
B4
B5
B6
B7
B8

Cluster

GFP
Pacific Blue
GFP+

%ChIP-seq Peaks

Distance to Nearest TSS (kb)

log2(normalized counts)

n=6 n=4 n=4 n=6

< -500
-500 to -50
-50 to -2
-2 to 0
> 500
50 to 500
2 to 50
0 to 2

0

50
Figure S5, Related to Figure 6.
Stone, Gifford, et al.

A

B

C

*\text*p < 1e-19

% Regions with Motif

\begin{tabular}{cccccc}
\text{Cluster} & C1 & C2 & C3 & C4 & C5 & C6 & C7 & C8 \\
\hline
\text{Gata} & * & * & * & * & * & & & \\
\text{Mef2} & * & * & * & * & * & & & \\
\text{Tbx} & * & * & * & * & * & & & \\
\end{tabular}

\begin{tabular}{cccccc}
\text{Mean Accessibility Fold Change} & C1 & C2 & C3 & C4 & C5 & C6 & C7 & C8 \\
\hline
\text{Gata4 (SF)} & & & & & & & & \\
\text{Mef2c (SF)} & & & & & & & & \\
\text{Tbx5 (SF)} & & & & & & & & \\
\text{GMT (AF)} & & & & & & & & \\
\end{tabular}
SUPPLEMENTAL FIGURE LEGENDS

Figure S1., Related to Figure 1.

(A) Representative fluorescence-activated cell sorting (FACS) plot for αMHC-GFP⁺ cells at week 2 of reprogramming.

(B) Expression [log₁₀(UMI+0.1)] of marker genes for each cluster visualized in UMAP plots. Numbered clusters from Figure 1B listed at right for reference.

Figure S2., Related to Figure 2.

(A) Expression [log₁₀(UMI+0.1)] of epicardial markers visualized in a UMAP trajectory plot.

(B) Violin plots representing normalized UMI levels of genes presented in Figures 2F-G for all clusters.

Figure S3., Related to Figure 3.

(A) Bar graph indicates total number of ATAC-seq peaks identified for each replicate (1-3). Sorted αMHC-GFP⁺ cells were assayed at each time point and in sorted dsRed-positive fibroblasts infected with a control dsRed retrovirus.

(B) Bar graph indicates total number of concordant ATAC-seq peaks for each condition and total number of replicate concordant peaks overlapping between conditions. Concordant peaks met an FDR threshold of < 0.05 in at least two out of three replicates.

(C) Principal component analysis of all ATAC-seq replicates (n=3) for each condition based on 10,000 most differentially accessible regions among all reprogramming time points compared with fibroblasts.

(D) Histogram shows distribution of regions within each dynamic cluster (A1-A8) and in a set of non-dynamic regions that are stably accessible in fibroblasts and in all reprogramming conditions, binned based on the distance from each peak to the nearest transcription start site (TSS). Dynamic regions were underrepresented within 2kb of a
TSS compared with non-dynamic regions accessible in all conditions (Fisher's p-value = 2.2e-16 for both up- and downstream bins).

(E) Heatmap displays expression levels for transcription factors with motifs with highest enrichment in dynamic accessibility clusters (A1-A8) compared with non-dynamic region set stably accessible in all conditions. Values indicate log2 of the mean normalized CPM across all replicates (n=3) from bulk RNA-seq completed on sorted αMHC-GFP+ cells collected at each time point during reprogramming.

**Figure S4., Related to Figure 5.**

(A) Representative fluorescence-activated cell sorting (FACS) plot for αMHC-GFP+ cells at day 2 of reprogramming, originating from immortalized neonatal cardiac fibroblasts.

(B) Heatmap displays tag densities of ChIP-seq sample and inputs, for all replicates. Counts were normalized for differences in sequencing depth between samples using upper quartile normalization separately for the ChIP and the input samples of each transcription factor. Columns represent independent replicates. Rows retain order shown in **Figure 5B**.

(C) Bar plot displays binned distribution of ChIP-seq peaks from nearest transcription start site (TSS). We identified a lower proportion of dynamic regions (all clusters) within 2kb up- and downstream of the nearest TSS compared with stably accessible, non-dynamic regions (Fisher odds ratio = 0.2 and 0.07, respectively; p-value = 2.2e-16 for both comparisons).

**Figure S5., Related to Figure 6.**

(A) Browser tracks display ChIP-seq in single factor and all factor conditions, and ATAC-seq signal in single factor, double factor, and all factor conditions, near leukocyte-endothelial cell adhesion molecule 2 (Lecam2). Track height was normalized to sample read depth within each assay. Orange rectangle highlights region from cluster 5 with Gata4 binding in the single factor condition which is ameliorated when all factors are present.
(B) Bar graph indicates percentage of regions within each cluster (C1-C8) that contain Gata4 (orange), Mef2c (blue), or Tbx5 (green) motifs. P-values were calculated using the cumulative hypergeometric distributions of motif occurrence in clustered region set compared with motif occurrence in non-dynamic regions stably accessible in fibroblasts and reprogramming conditions. Full motif enrichment results available in Table S7.

(C) Bar chart displays mean accessibility fold changes based on normalized tag densities in single infection (SF) conditions and in all factor (AF) condition, compared with fibroblasts, for regions within each cluster.

SUPPLEMENTAL TABLE LEGENDS

Table S1. Differential Gene Expression for Each Cluster, Related to Figure 1. Contains average log fold changes and p-values for all differentially expressed genes from negative binomial tests used to determine identities of populations in clusters in Figure 1B, and designates top 20 marker genes per cluster used to generate Figure 1E.

Table S2. Differential Gene Expression for Each Cluster, Related to Figure 2. Contains average log fold changes and p-values for all differentially expressed genes from negative binomial tests used to determine identities of populations in clusters in Figure 2A, and results of Moran’s I tests, including top two marker genes per cluster used in Figure 2C. Also contains average fold changes and p-values for all differentially expressed genes from pairwise tests of cluster 2 vs cluster 1, cluster 1 vs cluster 4, and cluster 4 vs cluster 8.

Table S3. Dynamic ATAC-seq Regions, Related to Figure 3. Table containing 100,691 regions dynamic between fibroblasts and αMHC-GFP+ cells collected at reprogramming time points day 2, day 3, week 1, week 2, and week 3. Second sheet contains clustered region identities from Figure 3B. Third sheet contains known motif enrichment results comparing regions in each cluster to stably accessible background region set, summarized in Figure 3D.
Table S4. Bulk RNA-seq Gene Expression during Reprogramming, Related to Figure 3 and Figure 4. Table includes normalized CPM values for each replicate over time course. Includes differential expression results between conditions.

Table S5. Predicted Co-regulatory Interactions of Transcription Factor Candidates, Related to Figure 4. Table includes total number of cofactors predicted by model for candidate regulators of reprogramming, with totals listed separately for predicted interactors with positive and negative net importance scores.

Table S6, Related to Figure 5. Table contains known motifs enriched in regions clustered by Gata4, Mef2c, Tbx5 occupancy and chromatin accessibility at day 2 of reprogramming, summarized in Figure 5C.

Table S7, Related to Figure 6. Known motifs enriched in regions clustered by Gata4, Mef2c, Tbx5 occupancy and chromatin accessibility during combinatorial exogenous expression of individual factors. Fold change in ChIP-seq signal over input and ATAC-seq signal over fibroblast control for all regions depicted in Figure 6C.
Supplemental Table 1

Click here to access/download

**Supplemental Videos and Spreadsheets**

Table.S1.Figure1.scRNAClusterDifferentialExpression.xl
Click here to access/download

**Supplemental Videos and Spreadsheets**

Table.S2.Figure2.scRNAClusterDifferentialExpression.xlsx
Click here to access/download
**Supplemental Videos and Spreadsheets**
Table.S3.Figure3.ATAC.DynamicRegions.Clusters.Motifs.xlsx
Click here to access/download
**Supplemental Videos and Spreadsheets**
Table.S4.BulkRNA_manuscript.only.csv
Click here to access/download
Supplemental Videos and Spreadsheets
Table.S6.Figure5_AF.ChIPCluster.Motifs.csv
SUPPLEMENTAL INFORMATION

Supplemental information includes 5 figures and 7 tables.