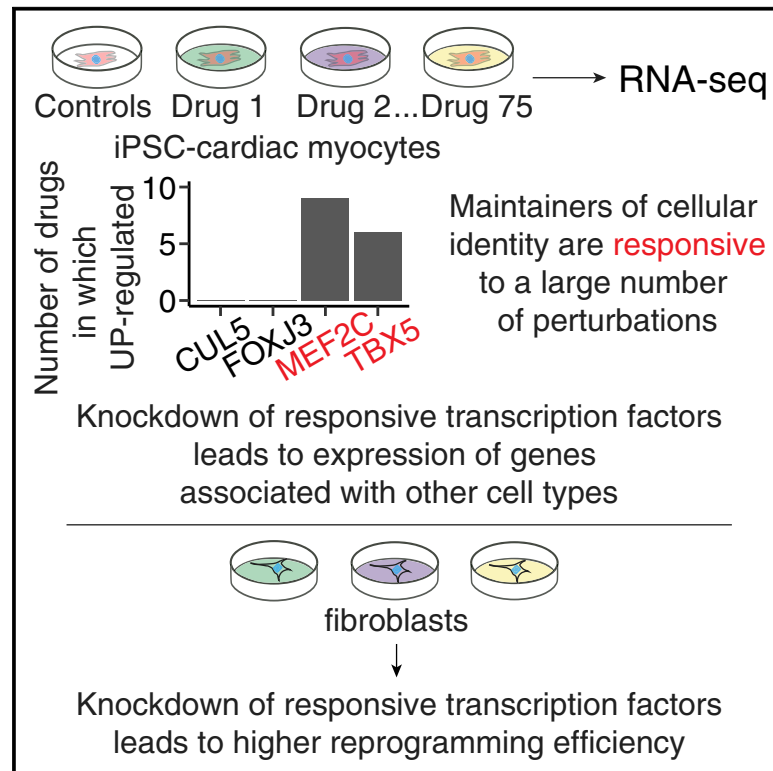# Responsiveness to perturbations is a hallmark of transcription factors that maintain cell identity *in vitro*

Graphical abstract

Authors

Ian A. Mellis, Hailey I. Edelstein, Rachel Truitt, ..., Wenli Yang, Rajan Jain, Arjun Raj

Correspondence

jainr@pennmedicine.upenn.edu (R.J.), arjunrajlab@gmail.com (A.R.)

In brief

Identifying transcription factors that maintain cell type is important for manipulating cell type. We hypothesized that because cell type is often resilient to perturbations, the transcriptional response to perturbations would reveal identity-maintaining genes. In human iPSC-derived cardiac myocytes, transcription factors important for cellular identity were among the most frequently upregulated (most responsive). Suppression of responsive transcription factors in fibroblasts enhanced reprogramming. We propose that responsiveness is a property of transcription factors that help maintain cellular identity.

Highlights

- Genes upregulated after many drug perturbations are "responsive" to perturbations

- Suppressing responsive transcription factors in iPSC-CMs upregulates nonmyocyte genes

- Suppressing responsive transcription factors in fibroblasts enhances reprogramming

- Responsive transcription factors often help maintain cell type

Article

Responsiveness to perturbations is a hallmark of transcription factors that maintain cell identity *in vitro*

Ian A. Mellis,[1,2] Hailey I. Edelstein,[3] Rachel Truitt,[3,4] Yogesh Goyal,[1,5] Lauren E. Beck,[1] Orsolya Symmons,[1] Margaret C. Dunagin,[1] Ricardo A. Linares Saldana,[6] Parisha P. Shah,[4,6] Juan A. Pérez-Bermejo,[7] Arun Padmanabhan,[8,9] Wenli Yang,[3,4,10] Rajan Jain,[3,4,6,10,11,*] and Arjun Raj[1,5,11,12,*]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA
[2]Genomics and Computational Biology Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[3]Institute for Regenerative Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[4]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[5]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[6]Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[7]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA
[8]Gladstone Institute of Cardiovascular Disease, San Francisco, CA, USA
[9]Division of Cardiology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA
[10]Penn Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[11]Penn Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[12]Lead contact
*Correspondence: jainr@pennmedicine.upenn.edu (R.J.), arjunrajlab@gmail.com (A.R.)
https://doi.org/10.1016/j.cels.2021.07.003

SUMMARY

Identifying the particular transcription factors that maintain cell type *in vitro* is important for manipulating cell type. Identifying such transcription factors by their cell-type-specific expression or their involvement in developmental regulation has had limited success. We hypothesized that because cell type is often resilient to perturbations, the transcriptional response to perturbations would reveal identity-maintaining transcription factors. We developed perturbation panel profiling (P³) as a framework for perturbing cells across many conditions and measuring gene expression responsiveness transcriptome-wide. In human iPSC-derived cardiac myocytes, P³ showed that transcription factors important for cardiac myocyte differentiation and maintenance were among the most frequently upregulated (most responsive). We reasoned that one function of responsive genes may be to maintain cellular identity. We identified responsive transcription factors in fibroblasts using P³ and found that suppressing their expression led to enhanced reprogramming. We propose that responsiveness to perturbations is a property of transcription factors that help maintain cellular identity *in vitro*. A record of this paper's transparent peer review process is included in the supplemental information.

INTRODUCTION

Cells of a particular cell type likely maintain many of their specific behavioral properties across a range of conditions. The maintenance of these properties can be the result of the activity of certain transcription factors, and the identification of these factors may be useful for manipulating cellular identity *in vitro* (Arendt et al., 2016). However, it has proven difficult to predict which transcription factors are responsible for the maintenance of the behaviors associated with that cellular identity and, therefore, which might be useful in converting cells from one type to another.

Transcription factors can regulate the phenotypic characteristics of a given cell type in a number of ways. In the context of

cell-type interconversion *in vitro*, two key functions of transcription factors are *identity-establishment* and *identity-maintenance*, meaning their ability to control the expression of genes that will establish or maintain the characteristics of a particular cell type, respectively. Here, we are interested in converting one cell type into another; hence, we define a factor that is "identity-establishing" as one whose addition leads to the acquisition of characteristics displayed by the target cell type (for instance, sarcomere formation in cardiac myocytes). We define "identity-maintaining" as one that decreases the ability of a cell to convert to another cell type (e.g., in transdifferentiation or reprogramming) *in vitro*. The distinction between cell-type transitions that occur *in vivo* versus *in vitro* is critical. *In vivo*, cell-type transitions of differentiated cells occur in normal biological processes
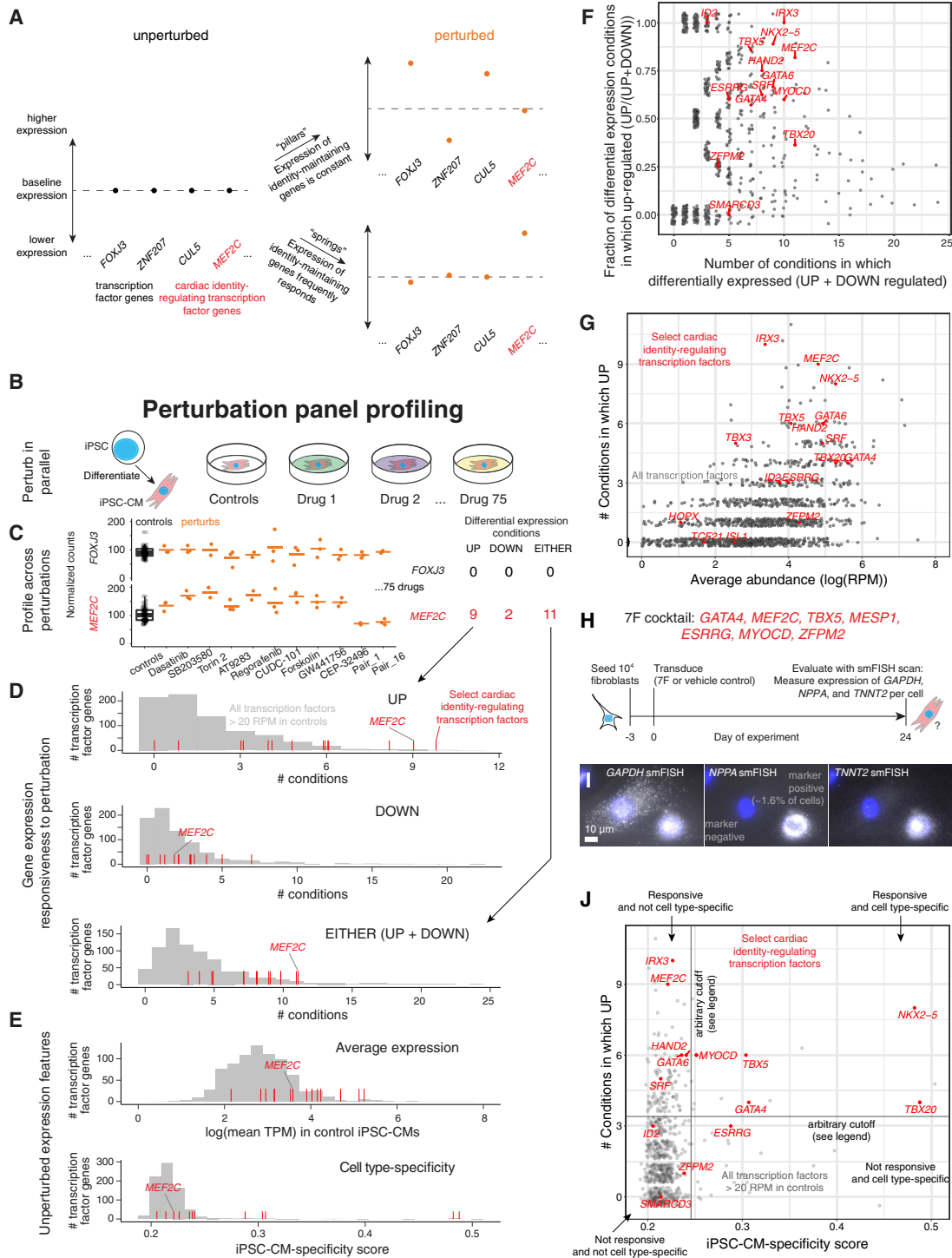
Figure 1. Perturbation panel profiling of iPSC-derived cardiac myocytes

(A) Conceptual guide for different perturbation-response models of identity-maintaining genes. In the "pillars" model, identity-maintaining genes are relatively unresponsive to perturbation, whereas non-identity-related genes do respond. In the "springs" model, identity-maintaining genes are more often responsive to perturbation than other genes.

(B) Perturbation panel profiling experiments and analysis: parallel cultures of human iPSC-derived cardiac myocytes (iPSC-CM), exposed to 75 different perturbation conditions in triplicate (see Table S1) or vehicle control.

(C) Parallel RNA extraction and library preparation for RNAtag-seq of perturbed and control samples followed by differential expression analysis. Per gene, differential expression in each condition relative to vehicle control samples followed by quantification of the number of conditions in which differential expression was observed. Representative results for two genes, *FOXJ3* and *MEF2C*, are shown.

(legend continued on next page)

ranging from regeneration to wound healing and in disease processes in which aberrant cell-type formation contributes, such as fibrosis, cancer, and other diseases. Several studies have used a variety of approaches to identify key regulators of such processes (e.g., loss of Nkx2-1 leading to loss of alveolar cell identity in lung adenocarcinoma [Snyder et al., 2013]) (Kikuchi et al., 2010; Jopling et al., 2010, 2011; Holmberg and Perlmann, 2012; Kheirollahi et al., 2019; Que et al., 2019). The ability to change cell types *in vitro*, which holds enormous practical potential, may, in principle, rely on distinct regulatory factors from those operating *in vivo* as the biological regulatory mechanisms are likely at least somewhat distinct (Jopling et al., 2011). Here, we have focused on identifying transcription factors that are important for cell-type interconversion in the *in vitro* setting, and the potential distinction from factors that operate *in vivo* make their identification challenging.

Many attempts to identify transcription factors that may be useful for identity maintenance have relied on cell-type specificity, meaning the degree to which a gene's expression is restricted to that one cell type and not to others. Part of the rationale for using cell-type-specific transcription factors is the fact that successful cell-type interconversion usually includes changes to the expression of genes that direct the functions of that cell type and suppression of sets of genes that direct the functions of other cell types (Ang et al., 2016; Cahan et al., 2014; Kathiriya et al., 2021; Morris et al., 2014). Cell-type specificity is a relatively easy property to measure, given the ubiquity of large gene expression datasets. Several recent approaches have mined these large gene expression datasets to try and identify functionally important transcription factors based on their cell-type specificity or inferred activity in particular tissues (Cahan et al., 2014; Chronis et al., 2017; D'Alessio et al., 2015; Morris et al., 2014; Rackham et al., 2016; Tomaru et al., 2014). The cell-type specificity of a transcription factor, however, fundamentally only correlates with the type. It is entirely possible that a transcription factor may be highly specific to a cell type but have no intrinsic ability to maintain the cell type, or that a transcription factor could be important for maintenance and yet not be particularly specific to any one cell type.

Another approach to identifying *in vitro* cellular-identity-maintaining transcription factors is to focus on transcription factors

known to drive the well-characterized processes of developmental lineage commitment or differentiation (Fu et al., 2013; Ieda et al., 2010; Leyva-Díaz and Hobert, 2019; Nam et al., 2013; Olson and Srivastava, 1996; Patel and Hobert, 2017; Qian et al., 2013; Rahe and Hobert, 2019; Takeuchi and Bruneau, 2009; Vierbuchen et al., 2010; Weintraub et al., 1989). Often, these studies start with a pool of factors and then test subsets of these factors for their ability to change the rate of reprogramming or transdifferentiation (Parekh et al., 2018; Takahashi and Yamanaka, 2006). However, the reprogramming or transdifferentiation of cells grown in culture is a fundamentally different process from development, and as such, there may be transcription factors that can maintain cell type *in vitro* that do not mirror normal development (Briggs et al., 2017; Mellis and Raj, 2015).

In light of these challenges, we hypothesized that the transcriptional response to perturbations would help identify genes whose expression is important for maintaining cell type. We envisioned two scenarios with distinct transcriptional response profiles to help maintain cellular identity. In the first scenario, critical transcription factors maintain stable expression in the face of perturbations, that is, they are the steady, non-responsive "pillars" of expression. In the second scenario, critical transcription factors are highly reactive and frequently respond to perturbations to ensure that cells maintain their type, that is, they are hyper-responsive "springs" of expression (Figure 1A). Transcriptionally profiling populations of cells subjected to a large panel of perturbations would allow us to discern whether one scenario was favored or if there was equipoise between the scenarios. Other perturbation profiling datasets have been generated, but their experimental designs have been geared toward other goals like drug target identification (Keenan et al., 2018; Niepel et al., 2019; Srivatsan et al., 2020; Szalai et al., 2019).

Here, we developed an experimental and analytical pipeline, perturbation panel profiling (P³), to measure the genome-wide transcriptional responses to a large set of perturbations in human iPSC-derived cardiac myocytes and fibroblasts. We demonstrated that well-known regulators of cardiogenesis are among the most responsive in their expression to a wide variety of perturbations in cardiac myocytes, i.e., they act as springs rather than pillars. High responsiveness was associated with transcription factors that suppressed the expression of

(D) Responsiveness to perturbation analysis of all expressed transcription factor genes. All transcription factor genes expressed at 20 RPM or greater in controls were considered (gray, histogram). Pre-registered set of cardiac myocyte identity-establishing and -maintaining genes shown in red. For each gene, the number of conditions in which it is up-regulated (top), down-regulated (middle), or differentially expressed either up or down (bottom) are shown.

(E) Static expression feature analysis of all expressed transcription factor genes. All transcription factor genes expressed at 20 RPM or greater in controls were considered (gray, histogram). iPSC-derived cardiac myocyte-specificity score is Jensen-Shannon specificity to iPSC-derived cardiac myocytes without GTEx Skin and Heart (see STAR Methods). Pre-registered set of cardiac myocyte identity-regulating genes shown in red. For each gene, the average expression level in vehicle control samples (top) or iPSC-derived cardiac myocyte-specificity score (bottom, see STAR Methods).

(F) Proportion of differential expression conditions in which upregulated versus the number of conditions in which up- or downregulated for transcription factor genes > 20 RPM in iPSC-CM controls. Known regulators of cardiac myocyte identity marked in red.

(G) Comparison of responsiveness (number of iPSC-CM conditions in which upregulated) and mean expression (mean RPM in control iPSC-CMs) for all expressed transcription factor genes. Vertical-only jitter added for visualization of all points with the same responsiveness value.

(H) Schematic of cardiac transdifferentiation experimental workflow. See STAR Methods for full details.

(I) Representative images of *GAPDH*, *NPPA*, and *TNNT2* single-molecule FISH (smFISH) of 7F-transduced fibroblasts after 24 days, demonstrating a cell that expresses marker genes *NPPA* and *TNNT2* at high levels and one cell that does not. Cells called "marker-positive" if *TNNT2* spots > 19 or *NPPA* spots > 24, thresholds were chosen arbitrarily. 0 cells in negative control MXs-empty condition are "marker-positive." Scale bars, 10 μm.

(J) Comparison of responsiveness and iPSC-derived cardiac myocyte-specificity scores for all expressed transcription factor genes. Vertical-only jitter added for visualization of all points with the same responsiveness value. Quadrant-separating lines ("arbitrary cutoffs") were chosen as elbows of the cumulative distribution function (cdf) of each feature overall transcription factor genes expressed at 20 RPM or greater in controls.

nonmyocyte genes in iPSC-derived cardiac myocytes. We then applied the P³ protocol to fibroblasts to identify highly responsive genes in that cell type. We showed that lowering the expression of responsive transcription factors identified by P³ in fibroblasts led to an enhancement in the cells' ability to reprogram into induced pluripotent stem cells. This enhancement suggested increased pliability of cell type, i.e., suppression of responsive transcription factors reduced fibroblast identity maintenance. In sum, our results demonstrate that responsiveness to perturbation may represent a property of transcription factor genes that are important for cell-type maintenance.

RESULTS

Transcription factor genes that regulate cellular identity are upregulated in cardiac myocytes after many different perturbations

To determine whether genes that are critical for cardiogenesis are more or less often responsive to perturbation than other similarly highly expressed genes, we designed an experimental and analytical workflow, P³. P³ consists of growing highly pure cell-type populations, exposing them to a panel of dozens of different drugs in parallel and performing RNAtag-sequencing (RNAtag-seq) for transcriptome-wide gene expression profiling in each condition. We then measure changes in gene expression across all conditions. We chose to initially apply P³ to human iPSC-derived cardiac myocytes (iPSC-CMs) because of the ease of culture, applicability to emerging therapies, and well-established developmental trajectories. To identify potential transcription factors in an unbiased way, we sought to perturb cells such that, in aggregate, we affected as much of the transcriptome as possible. Therefore, we sought to maximize the number of signaling cascades we perturbed, identifying 100 small-molecule drugs from the SelleckChem Bioactive library targeting any kinases or G protein-coupled receptors (GPCRs) and minimizing overlap in their annotated targets (Table S1). We administered these drugs individually or in pairs for a total of 75 different perturbation conditions (Figure S1A).

We cultured hundreds of parallel bulk samples of iPSC-derived cardiac myocytes, cultured under standard (i.e., control, n = 63 replicate samples, spread evenly across all plates) and different (N = 75) perturbed conditions (Figures 1B and S1A, n = 3 replicate samples per condition). The dose for each drug was chosen to be 100 times the median IC50 values (or 10 times for each drug in a pair) culled from the literature and product materials (we determined that 10×–100× were the appropriate dose factor empirically; data not shown; see STAR Methods), and ultimately applied each drug at one drug-specific dose in triplicate to samples that we further analyzed with RNAtag-seq (Shishkin et al., 2015). After treating iPSC-derived cardiac myocytes with each drug or DMSO-only exposure for 4 days, we found that 63 out of 75 conditions had wells with beating cells similar in morphology to vehicle-treated cells and without evident cell death (Figures S1B and S1C; Videos S1 and S2).

We next conducted transcriptome-wide gene expression profiling of the hundreds of perturbed (and control) samples of cardiac myocytes (Figure 1C). We performed RNAtag-seq in batches of 96 samples, distributing DMSO controls evenly across each batch. Gene expression profiles across the 458

quality-controlled samples (454 excluding 4 HeLa samples used as an outgroup) clustered first by cell type, and then often by perturbation condition (Figure S2). We identified hundreds to thousands (0–6,862 genes per condition) of differentially expressed genes in most perturbed conditions relative to corresponding cell-type controls (Figure S3A; STAR Methods). We sequenced our libraries to a depth at which we estimated that fully doubling the number of reads for all 458 samples would have increased the number of detected differentially expressed genes by at most only 48% for genes expressed at an average of 20 RPM or greater in controls (note that sequencing to half the depth would have reduced the number of detected differentially expressed genes by 29%; Figure S4).

In order to ensure our analysis was sufficiently unbiased, we verified that the majority of the coding genome was affected by a subset of our 60+ perturbations. We quantified the number of perturbations, which resulted in differential expression in iPSC-derived cardiac myocytes for each gene (Figure 1C), limiting our analyses to genes that were expressed at 20 RPM or higher in control iPSC-derived cardiac myocytes. We found that 94.5% of genes expressed at 20 RPM or greater in iPSC-derived cardiac myocytes were differentially expressed in at least 1 perturbation condition relative to the cell-type controls (Figure S3), demonstrating that the compounds in our library were able to perturb the vast majority of the transcriptome. We focused our subsequent analyses on transcription factors, as they control the expression of genes relevant to cellular identity and homeostasis.

Next, we determined the degree to which individual transcription factors were responsive in the iPSC-derived cardiac myocyte dataset. Our null hypothesis was that there is no correlation between transcription factor responsiveness and whether that transcription factor is known to regulate cardiac myocyte identity. We quantified the number of perturbations of iPSC-derived cardiac myocytes that resulted in differential expression (adjusted p value < 0.1) of a transcription factor versus controls. In addition, we wanted to ensure that responsiveness did not merely identify genes that were highly expressed or that only highly expressed genes were detectably responsive. We found that expression levels did not entirely correlate with detected responsiveness: many lowly expressed transcription factors were responsive to perturbations and many highly expressed transcription factors were not responsive to perturbation. However, we observed that multiple critical cardiac transcription factors frequently responded to perturbations by being upregulated, including *MEF2C, NKX2-5, GATA4, HAND2,* and *TBX5* (Später et al., 2014; Srivastava and Olson, 2000; Wang et al., 2002) (Figures 1D–1G). Furthermore, gene ontology analysis of all responsive genes (not just transcription factors) identified a clear over-enrichment for genes related to cardiac function (e.g., structural components such as sarcomeres and myofibrils) relative to all highly expressed genes. By contrast, non-responsive genes did not demonstrate this over-enrichment (Figure S6). One potential explanation for the enrichment of cardiac myocyte identity-regulating transcription factors in the set of responsive genes is that a subset of perturbation drugs may be accelerating the maturation of cultured cardiac myocytes. In order to test whether differential gene expression effects reflected further maturation instead of perturbation-response, we performed a

CellNet-based heart gene regulatory network activation score (heart score) assessment to measure similarity to human heart tissue samples (Radley et al., 2017). We found that both unperturbed and perturbed (23 perturbation conditions tested) had heart scores ranging from 0.912 to 0.969, with none of the perturbations showing any particular increase or decrease in heart score compared with control (for comparison, the heart score range of control fibroblasts was 0.0473 to 0.0700). Therefore, in all analyzed conditions, iPSC-derived cardiac myocytes heart scores are very high, while heart scores are low for fibroblasts. Also, within the set of different iPSC-derived cardiac myocyte conditions, we did not detect a correlation between heart scores and the number of upregulated transcription factors known to regulate cardiac myocyte identity (Figure S7). Thus, it seems unlikely that the perturbation results reflect accelerated maturation. Prominently among the responsive transcription factors were 6 of the 7 members of the "7F" cocktail, whose combined expression has previously been reported to induce transdifferentiation of fibroblasts into cardiac myocytes (Fu et al., 2013).

In order to show whether other measures of responsiveness showed similar enrichment for known regulators of cardiac myocyte identity as the above counting of conditions in which a gene is upregulated, we performed several additional analyses. First, we restricted responsiveness analysis to only consider $\log_2$ fold changes of magnitude greater than 0.5. Effect size-based filtering also showed enrichment of known regulators of cardiac myocyte cell identity among transcription factors upregulated more often than other similarly highly expressed transcription factors (Figures S5A and S5B). We also checked whether average differential expression effect sizes across the perturbation panel were higher for known regulators of cardiac myocyte identity than for other expressed transcription factors. We did not find any correlation between whether a transcription factor was important for cardiac myocyte identity and its average $\log_2$ fold change across all conditions. However, if we considered the absolute value of $\log_2$ fold changes, that is the magnitudes of differential expression effect sizes regardless of whether up- or downregulated, it did appear that known regulators of cardiac myocyte identity had higher mean absolute $\log_2$ fold changes than other similarly expressed transcription factors (Figures S5C and S5D). Second, we used Fisher's method of statistical meta-analysis for each gene to combine the p values of differential expression testing, as a way to check for consistency of a differential expression event across the entire perturbation panel of iPSC-derived cardiac myocytes for each gene. Meta-analysis showed that known regulators of cardiac myocyte cell identity tended to have lower combined p values than other similarly highly expressed transcription factors, which suggests they may be more consistently differentially expressed across the perturbation panel (Figure S5E; see STAR Methods). Third, we were curious to know whether high responsiveness at the level of transcription-factor-target sets (also known as regulons) also produced a set of transcription factors enriched for known regulators of cardiac myocyte cell identity. Therefore, we calculated regulon-level responsiveness scores for human heart tissue-based transcription factor regulons described in previous studies by combining the number of conditions in which each target gene is upregulated and the ratio of conditions in which up- to downregulated, normalized by the predicted strength of transcription-factor-target interaction (STAR Methods; Marbach et al., 2016). We found that regulons of transcription factors known to be regulators of cardiac myocyte cell identity tended to have higher regulon responsiveness scores than regulons of other highly expressed transcription factors (Figures S8A and S8B). Therefore, other measures of responsiveness to perturbations also suggest that regulators of cardiac myocyte cell identity tend to be differentially expressed more often than other similarly highly expressed transcription factors in iPSC-derived cardiac myocytes.

We next wanted to confirm that the 7F transcription factor cocktail could enable transdifferentiation in the specific fibroblast line we originally used to generate the iPSC-derived cardiac myocytes, so we overexpressed them in isogenic GM00942 fibroblasts by viral transduction. Using single-molecule RNA fluorescence in situ hybridization (FISH) (Raj et al., 2008), we found that around 1% of the cells expressed the cardiac myocyte-specific marker genes TNNT2 and NPPA. This efficiency is comparable with previously reported results (Fu et al., 2013; Mohamed et al., 2017). The appearance of these markers suggested that these cells were at least partially reprogrammed to a cardiac myocyte-like state (Figures 1H, 1I, and S9A–S9E), a finding we confirmed in two other fibroblast lines (Figure S9F). These transdifferentiation results confirm the importance of the 7F transcription factor cocktail in regulating cardiac myocyte identity. Notably, 6 of these 7 transcription factors were among the top hits in our P^3 experiments. Thus, transcription factors critical for the regulation of cardiac myocyte identity and cardiac function appear to be "springs," i.e., frequently responsive to perturbation, rather than "pillars" that hold steady in the face of perturbation. There are many reasons why these transcription factors might be more responsive to perturbations, including that they are part of a homeostatic response to perturbation or that they are simply downstream targets of many signaling pathways in the cells whose identity they regulate (see Discussion). One possible model for responsiveness is that transcription factors important for identity maintenance would frequently upregulate to help preserve identity in the face of a perturbation. A prediction of this model is that high responsiveness could signify that a transcription factor is important for the maintenance of cellular identity. In support of this model, several of these highly responsive transcription factors have been demonstrated to be identity-maintaining transcription factors, particularly NKX2-5, GATA4, and TBX5 (Ang et al., 2016; Kathiriya et al., 2021; Targoff et al., 2013). It is of course important to note that responsiveness may also correlate with a number of other properties that a transcription factor may have (Cui et al., 2018).

Responsiveness is a property of a gene's transcriptional regulation that appeared to be associated with identity-regulating transcription factors. Many approaches to date have, however, focused on cell-typespecificity, meaning whether the expression of a gene is specific to a given cell type or tissue type. To compare responsiveness to cell-type specificity, we used a cell-type-specificity score, which is a metric that combines expression levels and cell-type specificity as defined by various publicly available datasets. We calculated cell-type- (or tissue)-specificity scores for all genes in each tissue included in the GTEx dataset based on average expression levels (STAR Methods; Cabili et al., 2011; GTEx Consortium et al., 2017). In

order to compute transcription factor expression specificity in a way that would reflect the transcription factor expression levels in our own P^3 experiments, we replaced the heart and skin GTEx samples with expression data generated from our control iPSC-derived cardiac myocytes and GM00942 fibroblasts, respectively. We compared the responsiveness of transcription factors to their specificity scores for both iPSC-derived cardiac myocytes and GM00942 fibroblasts (Figures 1E and 1J). We found that responsiveness and cell-type-specificity score did not equate (similar results held when using exclusively GTEx samples without replacing the heart and skin samples [Figure S10]). Overall, we found that responsiveness was more sensitive than cell-type specificity for the identification of known factors: of the 14 select transcription factors known to be relevant to cardiac myocyte biology, cell-type-specificity score only identified 5 of them, whereas responsiveness identified 10. Cell-type specificity identified these factors far more specifically, though, in that those 5 identified transcription factors were in the top 24 factors ranked cell-type specificity, whereas responsiveness identified 10 transcription factors but from a pool of the top 133 ranked by responsiveness. It may be that these additional 123 factors are false positives in that they do not contribute to identity maintenance or unknown true positives in that they may in fact contribute to identity maintenance.

Knockdown of responsive transcription factors in iPSC-derived cardiac myocytes leads to upregulation of non-cardiac myocyte-associated genes

Previously published results showed that loss of GATA4 and TBX5, transcription factors known to regulate the development and long-term maintenance of essential cardiac myocyte functions, results in inappropriate upregulation of non-cardiac myocyte genes (Ang et al., 2016; Kathiriya et al., 2021). We, thus, reasoned that a way to test candidate transcription factors for a role in maintaining cellular identity was to knock down the transcription factor and look for the upregulation of gene expression programs important for other cell types. We selected 9 transcription factor genes from the four quadrants of the responsiveness versus specificity plot in Figure 1J, representing all combinations of responsive versus non-responsive and specific versus non-specific. Within these categories, there are several transcription factors that are known to be involved in cardiac development, along with several that were not. For the known transcription factors, we tested whether responsiveness was able to specifically identify ones that maintained identity, and for the unknown transcription factors, we tested whether responsiveness was able to identify novel and unexpected identity-maintaining transcription factors. Genes we tested that are known to be important for cardiac development include *NKX2.5* and *HAND1* (responsive and specific), *MEF2C* (responsive and not specific), *IRX4* (specific and not responsive), and *ZFPM2* (neither responsive nor specific). Genes we tested that are not known to be important for cardiac development include *SOX11* (responsive and specific), *HMGB2* (responsive and not specific), *SKIDA1* (specific and not responsive), and *COPS2* (neither responsive nor specific; see STAR Methods for the rationale for choosing candidates). We knocked down one gene at a time in WTC-11-derived iPSC-CMs using siRNA for 10 days, conducted bulk RNA-seq, and determined whether upregulated genes were enriched for

genes associated with other cell types (Figures 2A, 2B, and S11). We found that for the 9 tested genes, as we hypothesized, responsiveness was a strong predictor of whether a transcription factor was identity-maintaining compared with cell-type specificity. 4 of 5 tested responsive genes led to upregulation of neural, endothelial, or fibroblastic gene sets while only 1 of 4 tested unresponsive genes displayed this behavior. We did not find that iPSC-CM specificity was very predictive of whether knockdown of a transcription factor would lead to inappropriate upregulation of nonmyocyte gene sets (2 of 5 specific TFs and 3 of 4 nonspecific TFs showed non-cardiac myocyte term upregulation). Additionally, two knockdowns, MEF2C and ZFPM2, also led to upregulation of one myocyte-related gene set each, in addition to neural, endothelial, or fibroblastic gene sets. The 4 knockdowns that did not lead to upregulation of nonmyocyte gene sets also did not lead to upregulation of any other annotated GO-cellular component gene sets (Figure 2C). Therefore, for the 9 tested genes, a mix of transcription factors known to be important for cardiac development and unknown, we found that high responsiveness was a better predictor than the specificity of whether knockdown led to inappropriate upregulation of nonmyocyte gene sets. These results are consistent with our hypothesis that responsiveness is associated with whether a factor maintains cellular identity *in vitro*, and future higher-throughput studies will add to the generalizability of these findings.

Knockdown of putative identity-maintaining transcription factors identified by P^3 enhances reprogramming

The data generated from applying P^3 to iPSC-CMs suggested that genes that maintain identity *in vitro* are springs (frequently responsive to perturbations), as opposed to pillars. Specifically, the experiments demonstrated that responsiveness identified a cadre of genes known to, in part, regulate maintenance. Hence, we reasoned that responsiveness might also identify previously undescribed transcription factors that maintain cellular identity *in vitro* for other cell types of interest. One way to test this hypothesis is in reprogramming from a source to target cell type: we reasoned that the downregulation of transcription factors that are responsive in the source cell type may interfere with source-cell-type maintenance, manifesting as an enhancement of reprogramming to the target cell type. We tested this possibility by focusing on the induction of induced pluripotent stem cells from fibroblasts, i.e., fibroblast reprogramming.

To test the hypothesis that P^3 can reveal genes that, if depleted, would inhibit identity maintenance, we first performed P^3 on isogenic dermal fibroblasts (GM00942, Coriell, Figure 3A) to find responsive candidate transcription factor genes, which we could then evaluate for identity-maintaining behavior. Similar to our iPSC-derived cardiac myocyte studies, we found that 62 out of 75 fibroblast culture conditions had wells with cells with fibroblast morphology and without evident cell death. We also observed similar numbers of detectably differentially expressed genes in fibroblasts as in iPSC-derived cardiac myocytes across the perturbation panel (Figure S3). Responsive genes identified in our P^3 analysis again did not equate to highly expressed or cell-type-specific transcription factors in a parallel analysis to that described above for iPSC-CMs (Figures S12A, S12B, and S13). Further, Gene Ontology analysis on all responsive genes
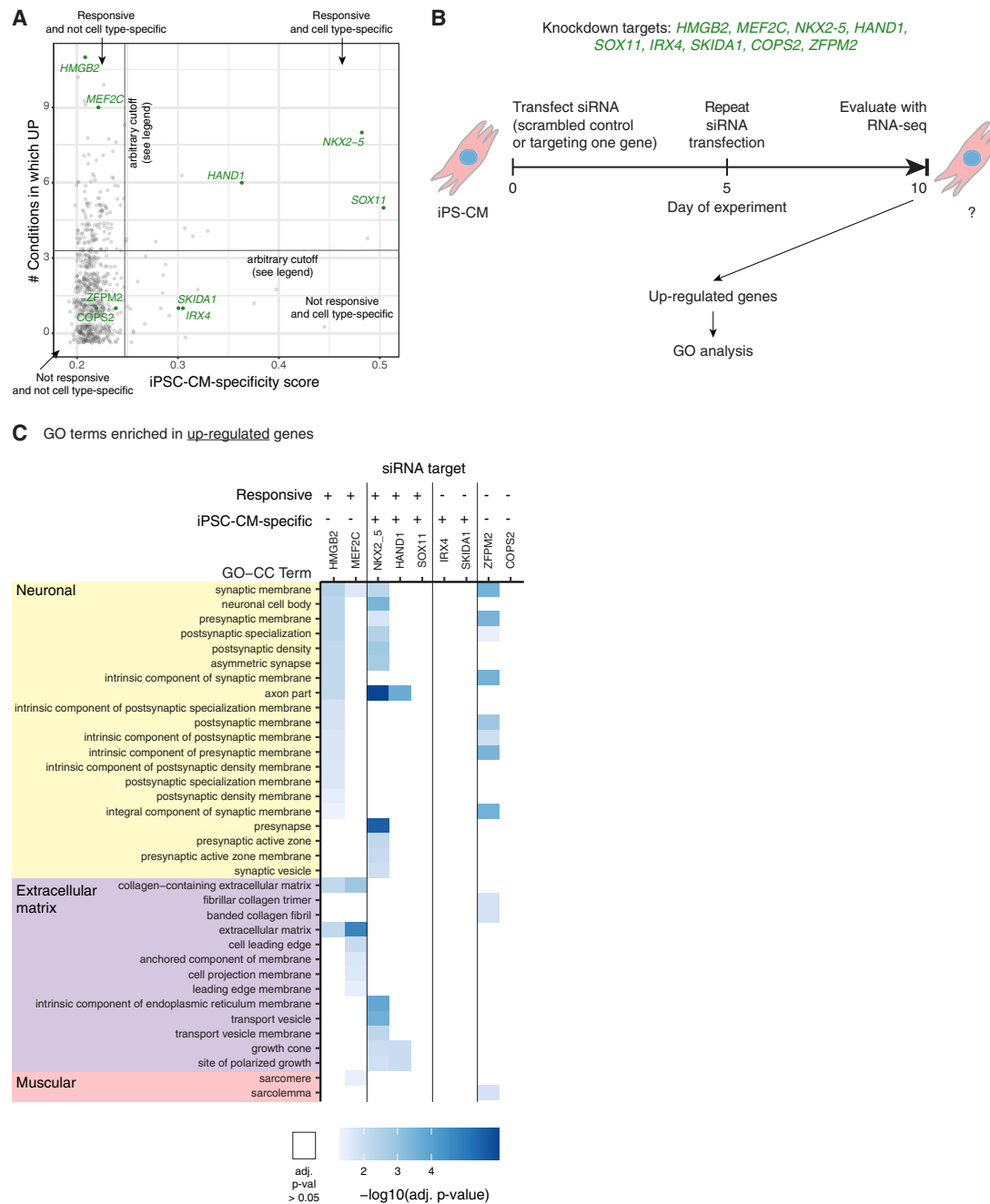
Figure 2. Suppression of transcription factor genes in iPSC-CMs

(A) Comparison of responsiveness and iPSC-derived cardiac myocyte-specificity scores for all expressed transcription factor genes, as in Figure 1I. Vertical-only jitter added for visualization of all points with the same responsiveness value. Quadrant-separating lines ("arbitrary cutoffs") were chosen as elbows of the cumulative distribution function of each feature overall transcription factor genes expressed at 20 RPM or greater in controls. Transcription factor genes chosen for knockdown are shown in green.

(B) Schematic of experimental workflow. Differentiated human iPSC-CMs (WTC-11) were transfected with small interfering RNA (siRNA) targeting a particular transcription factor, after which we performed RNA-seq.

(C) Significantly enriched GO-cellular component terms among upregulated genes for each knockdown. For upregulated genes: minimum $\log_2$-fold change magnitude = 0.5, maximum differential expression adjusted p value = 0.05. For GO terms: level = 4, maximum p value = 0.05, maximum q value = 0.05, simplified view.
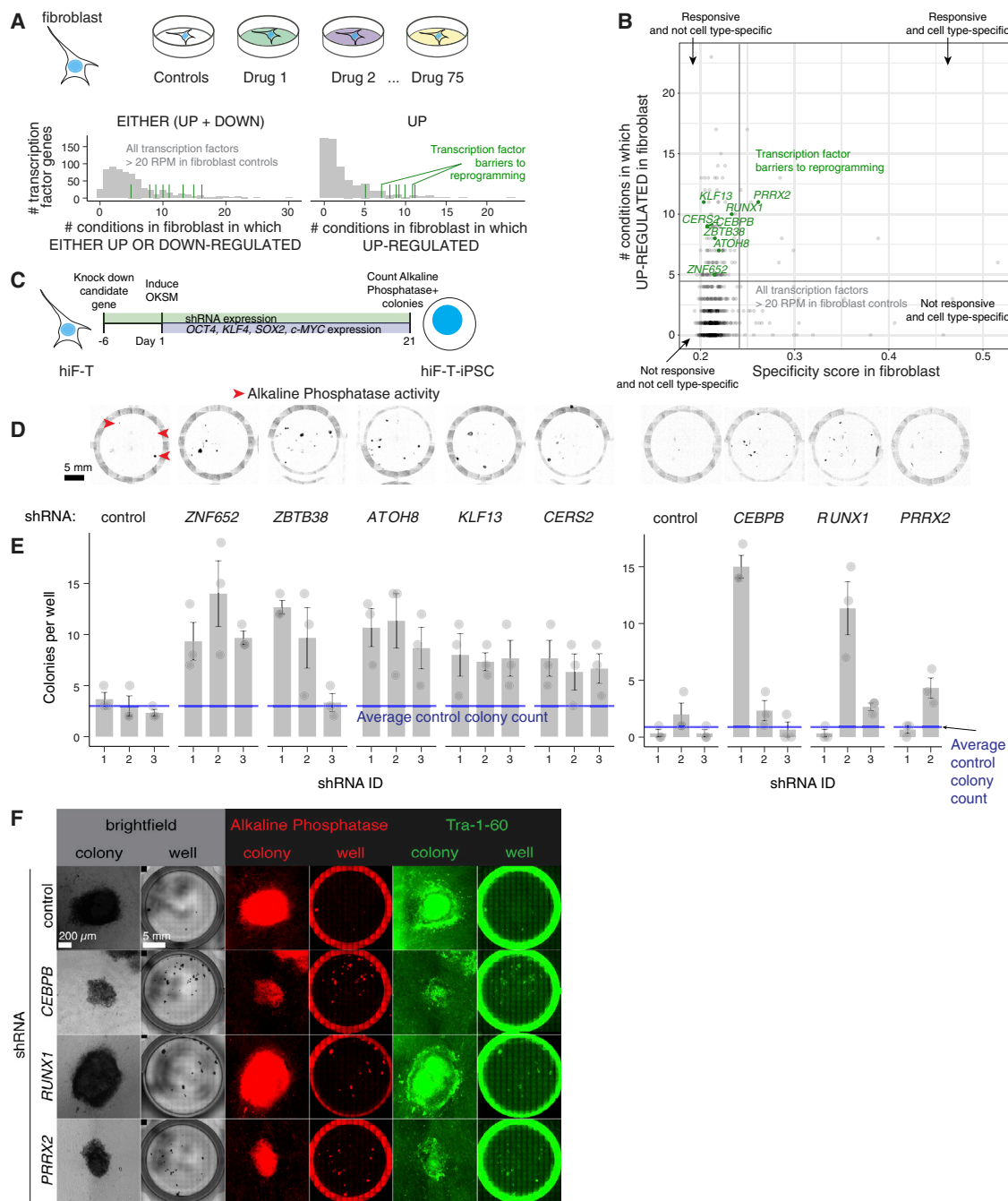
Figure 3. Suppression of fibroblast-responsive genes in a directed change of cell identity

(A) P³ schematic (top) and responsiveness results (bottom) for primary human GM00942 fibroblasts (primary fibroblasts) exposed to 75 different perturbation conditions and vehicle control. All transcription factors expressed at 20 RPM or greater in vehicle control fibroblasts (gray histogram). Selected responsive genes used in later reprogramming experiments are shown in green.

(B) Comparison of responsiveness and primary fibroblast-specificity score per gene (Jensen-Shannon specificity to primary fibroblasts without GTEx Heart and Skin, see STAR Methods). Quadrant-separating lines were chosen as elbows of the cdf of each feature over all transcription factor genes expressed at 20 RPM or greater in controls. Transcription factor barriers to reprogramming are 8 of 16 factors tested found to have a significant effect (Fisher's exact test p value < 0.05) on colony count against controls. See STAR Methods for a full list of factors tested.

(C) Schematic of experimental workflow for hiF-T fibroblast reprogramming experiments. In parallel, hiF-Ts were transduced with shRNAs targeting genes perturbable in fibroblasts (3 shRNAs per target, each cultured in technical triplicate per experimental replicate), expanded in culture for 6 days, counted, and seeded in parallel on MEF feeders. Dox-mediated OKSM induction for 20 days prior to fixation and analysis of reprogrammed colonies.

(legend continued on next page)

(not just transcription factors) revealed an over-enrichment for adhesion-related cellular components (e.g., cell-substrate junction) that are likely related to fibroblast function (Figure S6). This enrichment mirrors the enrichment of cardiac myocyte components among the set of responsive genes in iPSC-derived cardiac myocytes.

To test whether the set of responsive transcription factor genes was enriched for transcription factors whose downregulation could enhance reprogramming efficiency, we first chose 16 transcription factors that were often responsive in fibroblasts in our P[3] experiments, regardless of their cell-type specificity to fibroblasts. We considered any transcription factor genes that were (1) upregulated in at least 5 conditions in GM00942, (2) upregulated in at least 50% more conditions than they were downregulated in GM00942, (3) not also frequently upregulated in iPSC-derived cardiac myocytes, (4) expressed at >50 transcripts per million (TPM) in GM00942 controls, (5) not commonly studied in the context of fibroblast development based on literature review, (6) not commonly considered a member of a stress response or apoptosis pathway based on literature review, and (7) with at least 3 quality-controlled targeting shRNA clones available through our university core service lab's Human TRC 2.0 lentivirus library (see STAR Methods). Out of the set of 66 highly responsive transcription factors identified by P[3] in GM00942, 41 met all of the above criteria, of which we tested 16 for practical experimental scale considerations. We selected these 16 transcription factors for validation based on them having either the most conditions in which it was upregulated or the highest ratio of conditions in which it was up- versus downregulated. The human fibroblast line we used to test for reprogramming efficiency changes was the hiF-T line (Cacchiarelli et al., 2015), which has an integrated cassette encoding the Yamanaka factors (*OCT4, SOX2, KLF4, c-MYC*; OSKM Takahashi and Yamanaka, 2006) under doxycycline-inducible control. We used small hairpin RNAs (shRNAs) to knock down the expression of these 16 transcription factors one at a time in hiF-T cells prior to the induction of OSKM, then induced OSKM to check for differences in re-programmability. (Figures 3C and S14; Table S2). Of these 16 transcription factors, we found that 8 increased the number of alkaline-phosphatase-positive colonies after 3 weeks of OSKM expression (Figures 3B, 3D, 3E, S14, and S15). We further confirmed reprogramming by staining for Tra-1-60, a surface marker of induced pluripotent stem cells (Chan et al., 2009). We found that colonies positive for alkaline phosphatase activity also express Tra-1-60 (Figure 3F). Therefore, transcription factor gene expression responsiveness in fibroblasts identified genes important in maintaining fibroblast identity *in vitro*.

Next, we were curious whether high transcription-factor-target set (regulon) responsiveness in fibroblasts was predictive of whether knockdown of that factor led to increased reprogram-

ming frequency. Therefore, we calculated regulon-level responsiveness scores for human dermal fibroblast-based transcription factor regulons described in previous studies. Specifically, for each transcription factor, we combined the number of conditions in which each target gene is upregulated and the ratio of conditions in which up- to downregulated, normalized by the predicted strength of transcription-factor-target interaction (STAR Methods; Marbach et al., 2016). Only 5 of the 16 tested transcription factors have annotated regulons available for analysis, and all 5 were barriers to reprogramming. Therefore, we did not have enough data to test the hypothesis that regulon responsiveness is associated with whether a transcription factor is a barrier to reprogramming. However, for the few genes for which we did have data, we did not find that these 5 transcription factors all had high regulon-level responsiveness scores, only 2 of the 5 had high regulon-level responsiveness scores. Similarly, we did not find high regulon responsiveness scores for all 5 of the tested barriers to reprogramming when basing this analysis on related regulon definitions in normal skin fibroblasts or total skin tissue samples (Figures S8C–S8F). Therefore, we found that regulon-level responsiveness was not possible for all of the transcription factors that we found to be barriers to fibroblast reprogramming (whereas we could calculate gene-level responsiveness for all transcription factors), and that high regulon-level responsiveness was not predictive of whether a transcription factor would be a barrier to reprogramming in fibroblasts.

Lastly, we wondered whether responsiveness or identity-specificity could in general identify factors that were barriers to reprogramming (as compared to random sampling). We, thus, re-analyzed data from a published pooled shRNA screen for barriers to reprogramming in mouse embryonic fibroblasts (MEFs) (Borkent et al., 2016). We mapped mouse transcription factors to their human orthologs. Then, we compared each mouse gene's maximum post-reprogramming enrichment in the pooled screen against its human ortholog's responsiveness as calculated by P[3] (Figures 4A and 4B). We found that genes whose knockdown promoted reprogramming were roughly 3-times more likely to be identified as orthologs of responsive genes in our analysis (Figures 4C–4F). We also considered whether a similar association was held with identity-specific transcription factors, but there were relatively few orthologs of identity-specific transcription factors available to be tested, so our ability to quantitatively compare the relative predictive power of responsiveness versus type-specificity was limited (Figures 4G and 4H). Among the few fibroblast identity-specific genes, orthologs of specific transcription factor genes appeared to be barriers to reprogramming roughly a third more often than non-specific transcription factor genes (1/12 specific genes versus 8/132 non-specific genes), as well (Figure 4H). While our data are consistent with specificity being associated with being a barrier to reprogramming, the low

(D) Representative fluorescence microscopy images of alkaline phosphatase activity (Vector Red staining, 4× magnification, 555-nm wavelength channel) in reprogrammed hiF-Ts under different knockdown conditions after 20 days of OKSM induction. Activity is shown in black, as indicated by red arrowheads. Scale bar, 5 mm.

(E) Colony counts with alkaline phosphatase activity after reprogramming for representative experimental replicates for two batches of perturbable target genes. Error bars represent standard error of the mean.

(F) Further characterization of hiF-T-iPSC colonies after reprogramming under different knockdown conditions. 10× magnification brightfield images (top 2 rows), alkaline phosphatase activity by Vector Red (middle 2 rows), and Tra-1-60 expression as identified by immunofluorescence (bottom 2 rows) shown for the same wells and inset colonies. Scale bar, 5 mm for wells, 200 μm for colonies.
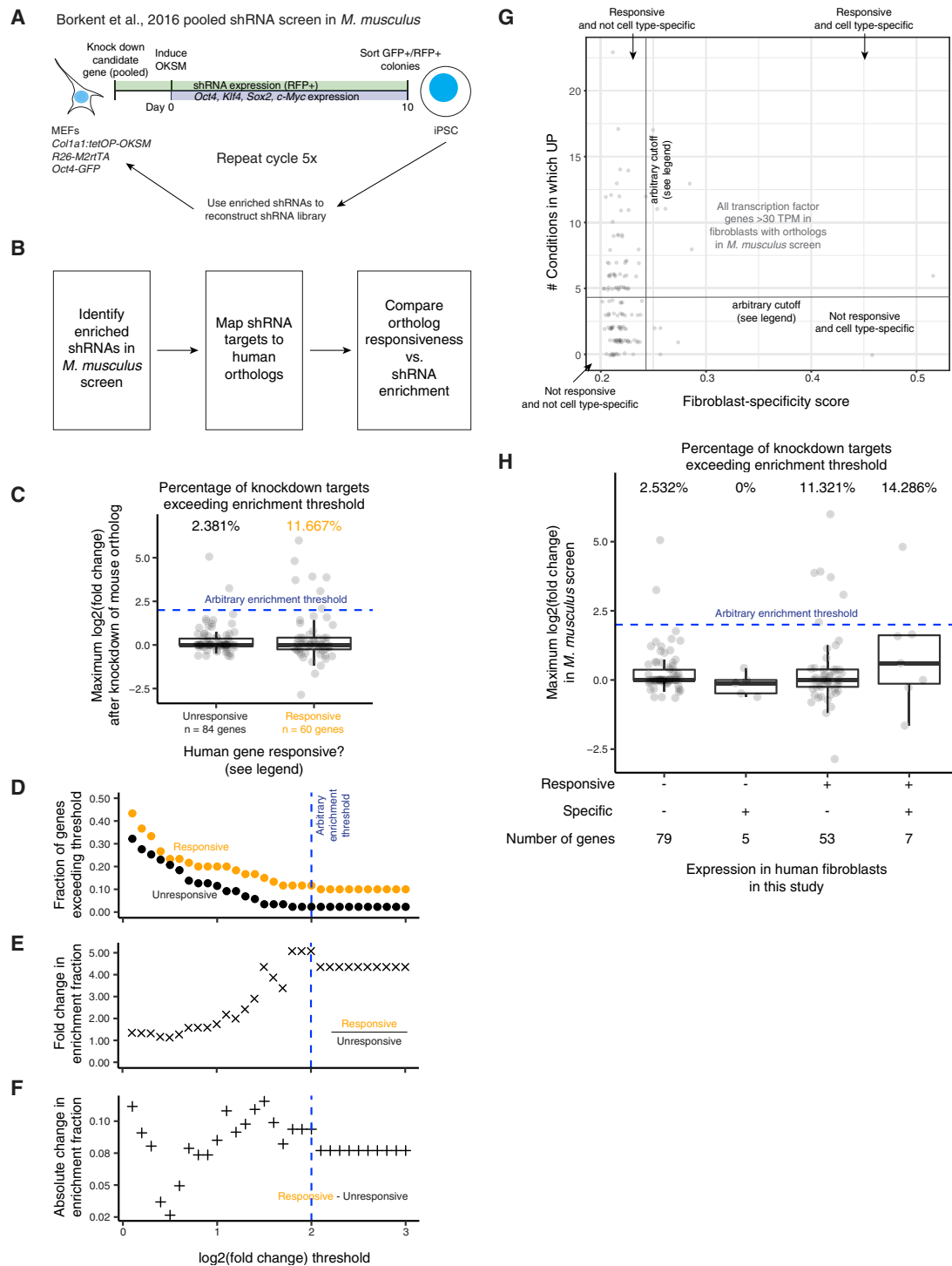
A Borkent et al., 2016 pooled shRNA screen in *M. musculus*

B

Identify enriched shRNAs in *M. musculus* screen → Map shRNA targets to human orthologs → Compare ortholog responsiveness vs. shRNA enrichment

C

D

E

F

G

H

Figure 4. Human ortholog responsiveness and enrichment in a pooled screen for barriers to mouse fibroblast reprogramming

(A) Experimental schematic of a published shRNA screen for barriers to mouse embryonic fibroblast reprogramming (Borkent et al., 2016).

(B) Analysis outline for comparing shRNA screen results (Borkent et al., 2016) with human responsiveness (this study).

(C) Maximum enrichment per gene across all targeting shRNAs in Borkent et al. (2016) screen for genes whose human orthologs are responsive (i.e., number of conditions in which upregulated > 4) or unresponsive. Considering only transcription factor genes expressed > 30 TPM in controls (arbitrary cutoff to consider only the highest-expressed transcription factors).

(D) Fraction of responsive and unresponsive gene orthologs exceeding different shRNA screen enrichment cutoffs.

(E) Fold change in a fraction of orthologs exceeding different shRNA screen enrichment cutoffs when considering responsiveness.

(F) Absolute difference in the fraction of orthologs exceeding different shRNA screen enrichment cutoffs when considering responsiveness.

(legend continued on next page)

number of genes available for comparison prevent us from making a strong claim in that regard.

DISCUSSION

The ubiquity of transcriptomic measurements has enabled us to profile the expression levels of genes across virtually all tissue types and, soon enough, all cell types. However, while these expression profiles may reveal which transcription factors are associated with a particular set of cellular functions, such transcription factors may or may not play a functional role in maintaining said functions. We have found that high responsiveness of expression to perturbations may be a way to identify transcription factors that play this functional role, i.e., these transcription factors are "springs" that pull cells back to their identity, rather than "pillars" that remain constant in the face of perturbations.

One of the most prevalent ways to identify candidate transcription factors for maintaining cell type is to analyze compendiums of transcriptional profiles across various tissue types and look for genes whose expression or activity is cell-type- or tissue-specific. In principle, responsiveness could have identified a proper subset of the cell-type-specific genes, a superset of the cell-type-specific genes, a wholly distinct set of genes, or a partially overlapping subset. We found that responsiveness seemed to identify a superset of the cell-type-specific genes, meaning that there were many genes whose expression was responsive but not particularly cell-type-specific. When we explicitly tested whether these responsive but not cell-type-specific genes could affect maintenance, we found that, for fully half the transcription factors we tested, suppression of the transcription factor led to an increase in reprogramming efficiency. This validation rate suggests that these responsive transcription factors are indeed important for cell-type maintenance and not merely falsely identified by P^3. The fact that these non-cell-type-specific transcription factors are important for maintenance suggests that cell-type maintenance may involve transcription factors that act as general barriers to cell-type transformations in addition to transcription factors specific to particular types (although it is possible that such transcription factors may only act as barriers in particular cell types). Our results indicate that there are many more cell-type maintenance transcription factors than cell-type specificity alone would suggest.

It is important to note here that cellular identity maintenance may be only one of many important roles that are associated with responsive transcription factors. Our "springs" model predicts that transcription factors that maintain identity are much more likely to be responsive to perturbations. However, there are a large number of potentially overlapping functions that any given transcription factor can have, such as identity-establishment, regulation of cell-type-specific genes, and several others. Determining which of these properties most strongly associates with responsiveness would require extensive functional testing of all such properties together with P^3 (Crow et al., 2019). It would

also be interesting to extend the analysis of responsiveness by P^3 beyond just transcription factors. Other modes of regulation, such as DNA methylation and chromatin remodeling, may be critical for behaviors such as identity maintenance. However, we chose to focus on transcription factors for our study given their widespread use in reprogramming protocols (Morez et al., 2015; Rahe and Hobert, 2019; Rulands et al., 2018). It will be interesting to see whether the responsiveness of factors responsible for other forms of regulation will also help identify the ones important for identity maintenance, or whether they have different responsiveness profiles.

We posit there may be at least two reasons why identity-maintaining transcription factors may be highly responsive to perturbation. One possibility is that identity-maintaining transcription factors are simply downstream of more signaling pathways than other transcription factors in the cell type that they maintain, hence their increased frequency of responsiveness to a panel of perturbations. However, given that the responsiveness of genes was different in the two cell types that we performed P^3 in, such a rationale would require the often invoked "cell-type-specific regulatory differences". Another possibility is that the upregulation of these genes is part of a feedback mechanism that mediates a homeostatic response to a perturbation. Such a response may be particular to specific signals or may be part of a more general response to a perturbative change in some aspect of the physiology or regulation of a cell's identity. Given that the genes identified by P^3 were responsive to several perturbations, we suspect the latter. Another interesting question about responsiveness is whether the responses are transient or sustained. The presence of negative feedback loops or other homeostatic mechanisms may temper the signal over time, leading to transient responses, whereas the lack of such mechanisms may yield sustained responses. Temporal measurements may help distinguish between transient and stable responses.

It is important to distinguish between identity-establishing transcription factors versus identity-maintaining transcription factors. Although some identity-establishing genes are also identity-maintaining, such as GATA4 and TBX5 in cardiac myocytes, OCT4 in pluripotent stem cells, and MYT1L in neurons (Ang et al., 2016; Mall et al., 2017; Shi and Jin, 2010), the converse need not be frequently true. That is, under the interpretation that responsiveness is a hallmark of identity maintenance, one could assume that responsive transcription factors would not necessarily also help establish a new cellular identity altogether. Consistent with this logic, we found that the transdifferentiation of fibroblasts to cardiac myocytes was not enhanced by adding responsive transcription factors to the existing seven transcription factor (7F) cocktail (Figure S16) (technically, however, it is far more difficult to evaluate the effects of overexpressing large combinations of transcription factors [Addis et al., 2013; Guo and Morris, 2017; Nam et al., 2014; Qian et al., 2013; Wang et al., 2015]). Other approaches may be required to solve that complementary question.

(G) Comparison of responsiveness and primary fibroblast-specificity score per gene (Jensen-Shannon specificity to primary fibroblasts without GTEx Heart and Skin, see STAR Methods), considering only transcription factor genes > 30 TPM in control fibroblasts with mouse orthologs tested in Borkent et al. (2016) screen. Same cutoffs as in Figure 3B.

(H) Maximum enrichment per gene across all targeting shRNAs in Borkent et al. (2016) screen for genes whose human orthologs are responsive (i.e., number of conditions in which upregulated > 4) or unresponsive, specific (score > 0.24) or non-specific.

An alternative approach that has been used to find cellular-identity-maintaining transcription factors is genetic or small-molecule screens, often based on limited libraries of candidates that direct differentiation *in vivo* (Cacchiarelli et al., 2015; Ebrahimi et al., 2019; Liu et al., 2018; Parekh et al., 2018; Zhou et al., 2017). Cacchiarelli et al. used pooled shRNA screens targeted toward a set of epigenetic factors to identify a number of such maintenance factors. It is difficult, however, to perform such screens in a genome-wide fashion owing to technical considerations. Other groups have screened small molecules that may affect reprogramming, and one such effort discovered the importance of *PRRX1*, a family member of *PRRX2*, the latter of which we found to maintain fibroblast identity (Ebrahimi et al., 2019). We envision that P^3 may serve as a useful complement to such screens by providing a means by which one could identify on the order of a hundred candidate transcription factors, which could subsequently be tested in a pooled screen. Such a combined approach may allow for rapid and unbiased identification of identity-maintaining transcription factors.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Cell culture
  - GM00942 human dermal fibroblast culture
  - GM11169 human cardiac fibroblast culture
  - HEK293FT culture
  - hiF-T culture
  - Immortalized human cardiac fibroblast (immHCF) culture
  - Platinum-A (Plat-A) retroviral packaging cell line culture
  - Cellular reprogramming
  - Perturbation culture
  - Transdifferentiation of fibroblasts to induced cardiac myocyte-like cells
  - Cloning of transcription factor genes and TurboGFP into pMXs
  - Titering of pMXs retroviral vectors
  - Gene knockdown in iPSC-derived cardiomyocytes
  - shRNA-mediated knockdown of transcription factors in hiF-T cells
  - Verification of knockdown efficiency following shRNA transduction
  - hiF-T reprogramming to pluripotency
  - High-throughput RNA extraction
  - RNAtag sequencing
  - RNA sequencing
  - RNAtag-seq and RNA-seq data processing
  - Live cell Tra-1-60 imaging
  - Alkaline phosphatase staining with colorimetry
  - Immunofluorescence
  - Single-molecule RNA FISH
  - Imaging
  - Image Processing
  - Reproducible analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gene expression perturbation-responsiveness
  - Gene expression cell type-specificity analysis
  - Gene set enrichment analysis
  - Meta-analysis of differential expression
  - Transcription factor regulon responsiveness analysis
  - Prioritization of highly responsive genes for use in reprogramming experiments

AUTHOR CONTRIBUTIONS

A.R. and I.A.M. conceived of P^3, and I.A.M., R.J., and A.R. conceived of the larger project with applications to cardiac transdifferentiation and reprogramming. I.A.M. designed, conducted, and analyzed all experiments with supervision by A.R. and R.J., with the exception of iPSC-CM differentiation performed by H.I.E., R.T., and W.Y. O.S. contributed to the design of some

Cell Systems
Article

REFERENCES

Addis, R.C., Ifkovits, J.L., Pinto, F., Kellam, L.D., Esteso, P., Rentschler, S., Christoforou, N., Epstein, J.A., and Gearhart, J.D. (2013). Optimization of direct fibroblast reprogramming to cardiomyocytes using calcium activity as a functional measure of success. J. Mol. Cell. Cardiol. 60, 97–106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Ang, Y.S., Rivas, R.N., Ribeiro, A.J.S., Srivas, R., Rivera, J., Stone, N.R., Pratt, K., Mohamed, T.M.A., Fu, J.-D., Spencer, C.I., et al. (2016). Disease model of GATA4 mutation reveals transcription factor cooperativity in human cardiogenesis. Cell 167, 1734–1749.e22.

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., and Wagner, G.P. (2016). The origin and evolution of cell types. Nat. Rev. Genet. 17, 744–757.

Borkent, M., Bennett, B.D., Lackford, B., Bar-Nur, O., Brumbaugh, J., Wang, L., Du, Y., Fargo, D.C., Apostolou, E., Cheloufi, S., et al. (2016). A serial shRNA screen for roadblocks to reprogramming identifies the protein modifier SUMO2. Stem Cell Rep. 6, 704–716.

Briggs, J.A., Li, V.C., Lee, S., Woolf, C.J., Klein, A., and Kirschner, M.W. (2017). Mouse embryonic stem cells can differentiate via multiple paths to the same state. eLife 6, e26945.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915–1927.

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

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell 158, 903–915.

Chan, E.M., Ratanasirintrawoot, S., Park, I.H., Manos, P.D., Loh, Y.H., Huo, H., Miller, J.D., Hartung, O., Rho, J., Ince, T.A., et al. (2009). Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. Nat. Biotechnol. 27, 1033–1037.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. Cell 168, 442–459.e20.

Crow, M., Lim, N., Ballouz, S., Pavlidis, P., and Gillis, J. (2019). Predictability of human differential gene expression. Proc. Natl. Acad. Sci. USA 116, 6491–6500.

Cui, M., Wang, Z., Bassel-Duby, R., and Olson, E.N. (2018). Genetic and epigenetic regulation of cardiomyocytes in development, regeneration and disease. Development 145, dev171983.

D'Alessio, A.C., Fan, Z.P., Wert, K.J., Baranov, P., Cohen, M.A., Saini, J.S., Cohick, E., Charniga, C., Dadon, D., Hannett, N.M., et al. (2015). A systematic approach to identify candidate transcription factors that control cell identity. Stem Cell Rep. 5, 763–775.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Ebrahimi, A., Sevinç, K., Gürhan Sevinç, G., Cribbs, A.P., Philpott, M., Uyulur, F., Morova, T., Dunford, J.E., Göklemez, S., Arı, Ş., et al. (2019). Bromodomain inhibition of the coactivators CBP/EP300 facilitate cellular reprogramming. Nat. Chem. Biol. 15, 519–528.

Fu, J.D., Stone, N.R., Liu, L., Spencer, C.I., Qian, L., Hayashi, Y., Delgado-Olguin, P., Ding, S., Bruneau, B.G., and Srivastava, D. (2013). Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. Stem Cell Rep. 1, 235–247.

GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on gene expression across human tissues. Nature 550, 204–213.

Guo, C., and Morris, S.A. (2017). Engineering cell identity: establishing new gene regulatory and chromatin landscapes. Curr. Opin. Genet. Dev. 46, 50–57.

Holmberg, J., and Perlmann, T. (2012). Maintaining differentiated cellular identity. Nat. Rev. Genet. 13, 429–439.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature 460, 1132–1135.

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

Jopling, C., Boue, S., and Izpisua Belmonte, J.C. (2011). Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nat. Rev. Mol. Cell Biol. 12, 79–89.

Jopling, C., Sleep, E., Raya, M., Martí, M., Raya, A., and Izpisúa Belmonte, J.C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. Nature 464, 606–609.

Judge, L.M., Perez-Bermejo, J.A., Truong, A., Ribeiro, A.J., Yoo, J.C., Jensen, C.L., Mandegar, M.A., Huebsch, N., Kaake, R.M., So, P.-L., et al. (2017). A BAG3 chaperone complex maintains cardiomyocyte function during proteotoxic stress. JCI Insight 2, e94623.

Kathiriya, I.S., Rao, K.S., Iacono, G., Devine, W.P., Blair, A.P., Hota, S.K., Lai, M.H., Garay, B.I., Thomas, R., Gong, H.Z., et al. (2021). Modeling human TBX5 haploinsufficiency predicts regulatory networks for congenital heart disease. Dev. Cell 56, 292–309.e9.

Keenan, A.B., Jenkins, S.L., Jagodnik, K.M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A.B., Silverstein, M.C., Lachmann, A., et al. (2018). The library of Integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. Cell Syst. 6, 13–24.

Kheirollahi, V., Wasnick, R.M., Biasin, V., Vazquez-Armendariz, A.I., Chu, X., Moiseenko, A., Weiss, A., Wilhelm, J., Zhang, J.S., Kwapiszewska, G., et al. (2019). Metformin induces lipogenic differentiation in myofibroblasts to reverse lung fibrosis. Nat. Commun. 10, 2987.

Kikuchi, K., Holdway, J.E., Werdich, A.A., Anderson, R.M., Fang, Y., Egnaczyk, G.F., Evans, T., Macrae, C.A., Stainier, D.Y.R., and Poss, K.D. (2010). Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. Nature 464, 601–605.

Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S., et al. (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nat. Biotechnol. 25, 1015–1024.

Leyva-Díaz, E., and Hobert, O. (2019). Transcription factor autoregulation is required for acquisition and maintenance of neuronal identity. Development 146, dev177378.

Liu, Y., Yu, C., Daley, T.P., Wang, F., Cao, W.S., Bhate, S., Lin, X., Still, C., 2nd, Liu, H., Zhao, D., et al. (2018). CRISPR activation screens systematically identify factors that drive neuronal fate and reprogramming. Cell Stem Cell 23, 758–771.e8.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Maddah, M., Heidmann, J.D., Mandegar, M.A., Walker, C.D., Bolouki, S., Conklin, B.R., and Loewke, K.E. (2015). A non-invasive platform for functional characterization of stem-cell-derived cardiomyocytes with applications in cardiotoxicity testing. Stem Cell Rep. 4, 621–631.

Mall, M., Kareta, M.S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S.D., Ge, X., Drake, S., Euong Ang, C., et al. (2017). Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. Nature 544, 245–249.

Marbach, D., Lamparter, D., Quon, G., Kellis, M., Kutalik, Z., and Bergmann, S. (2016). Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat. Methods 13, 366–370.

Mellis, I.A., and Raj, A. (2015). Half dozen of one, six billion of the other: what can small- and large-scale molecular systems biology learn from one another? Genome Res. 25, 1466–1472.

Miyaoka, Y., Chan, A.H., Judge, L.M., Yoo, J., Huang, M., Nguyen, T.D., Lizarraga, P.P., So, P.L., and Conklin, B.R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. Nat. Methods 11, 291–293.

Mohamed, T.M.A., Stone, N.R., Berry, E.C., Radzinsky, E., Huang, Y., Pratt, K., Ang, Y.S., Yu, P., Wang, H., Tang, S., et al. (2017). Chemical enhancement of in vitro and in vivo direct cardiac reprogramming. Circulation 135, 978–995.

Morez, C., Noseda, M., Abreu Paiva, M., Belian, E., Schneider, M.D., and Stevens, M.M. (2015). Enhanced efficiency of genetic programming toward cardiomyocyte creation through topographical cues. Biomaterials 70, 94–104.

Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K.S., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell 158, 889–902.

Nam, Y.J., Lubczyk, C., Bhakta, M., Zang, T., Fernandez-Perez, A., McAnally, J., Bassel-Duby, R., Olson, E.N., and Munshi, N.V. (2014). Induction of diverse cardiac cell types by reprogramming fibroblasts with cardiac transcription factors. Development 141, 4267–4278.

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

Niepel, M., Hafner, M., Mills, C.E., Subramanian, K., Williams, E.H., Chung, M., Gaudio, B., Barrette, A.M., Stern, A.D., Hu, B., et al. (2019). A multi-center study on the reproducibility of drug-response assays in mammalian cell lines. Cell Syst. 9, 35–48.e5.

Olson, E.N., and Srivastava, D. (1996). Molecular pathways controlling heart development. Science 272, 671–676.

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., and Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol. Cell 58, 339–352.

Palpant, N.J., Pabon, L., Roberts, M., Hadland, B., Jones, D., Jones, C., Moon, R.T., Ruzzo, W.L., Bernstein, I., Zheng, Y., and Murry, C.E. (2015). Inhibition of β-catenin signaling respecifies anterior-like endothelium into beating human cardiomyocytes. Development 142, 3198–3209.

Parekh, U., Wu, Y., Zhao, D., Worlikar, A., Shah, N., Zhang, K., and Mali, P. (2018). Mapping cellular reprogramming via pooled overexpression screens with paired fitness and single-cell RNA-sequencing readout. Cell Syst. 7, 548–555.e8.

Patel, T., and Hobert, O. (2017). Coordinated control of terminal differentiation and restriction of cellular plasticity. eLife 6, e24100.

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

Que, J., Garman, K.S., Souza, R.F., and Spechler, S.J. (2019). Pathogenesis and cells of origin of Barrett's esophagus. Gastroenterology 157, 349–364.e1.

Rackham, O.J.L., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., FANTOM Consortium, Suzuki, H., Nefzger, C.M., Daub, C.O., Daub, C.O., et al. (2016). A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. 48, 331–335.

Radley, A.H., Schwab, R.M., Tan, Y., Kim, J., Lo, E.K.W., and Cahan, P. (2017). Assessment of engineered cells using CellNet and RNA-seq. Nat Protoc 12, 1089–1102.

Rahe, D.P., and Hobert, O. (2019). Restriction of cellular plasticity of differentiated cells mediated by chromatin modifiers, transcription factors and protein kinases. G3 (Bethesda) 9, 2287–2302.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods 5, 877–879.

Rulands, S., Lee, H.J., Clark, S.J., Angermueller, C., Smallwood, S.A., Krueger, F., Mohammed, H., Dean, W., Nichols, J., Rugg-Gunn, P., et al. (2018). Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. Cell Syst 7, 63–76.

Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature 546, 431–435.

Shanmughapriya, S., Tomar, D., Dong, Z., Slovik, K.J., Nemani, N., Natarajaseenivasan, K., Carvalho, E., Lu, C., Corrigan, K., Garikipati, V.N.S., et al. (2018). FOXD1-dependent MICU1 expression regulates mitochondrial activity and cell differentiation. Nat. Commun. 9, 3449.

Shi, G., and Jin, Y. (2010). Role of Oct4 in maintaining and regaining stem cell pluripotency. Stem Cell Res. Ther. 1, 39.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nat. Methods 12, 323–325.

Snyder, E.L., Watanabe, H., Magendantz, M., Hoersch, S., Chen, T.A., Wang, D.G., Crowley, D., Whittaker, C.A., Meyerson, M., Kimura, S., and Jacks, T. (2013). Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. Mol. Cell 50, 185–199.

Später, D., Hansson, E.M., Zangi, L., and Chien, K.R. (2014). How to make a cardiomyocyte. Development 141, 4418–4431.

Srivastava, D., and Olson, E.N. (2000). A genetic blueprint for cardiac development. Nature 407, 221–226.

Srivatsan, S.R., McFaline-Figueroa, J.L., Ramani, V., Saunders, L., Cao, J., Packer, J., Pliner, H.A., Jackson, D.L., Daza, R.M., Christiansen, L., et al. (2020). Massively multiplex chemical transcriptomics at single-cell resolution. Science 367, 45–51.

Szalai, B., Subramanian, V., Holland, C.H., Alföldi, R., Puskás, L.G., and Saez-Rodriguez, J. (2019). Signatures of cell death and proliferation in perturbation transcriptomics data-from confounding factor to effective prediction. Nucleic Acids Res 47, 10010–10026.

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

Takeuchi, J.K., and Bruneau, B.G. (2009). Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. Nature 459, 708–711.

Targoff, K.L., Colombo, S., George, V., Schell, T., Kim, S.H., Solnica-Krezel, L., and Yelon, D. (2013). Nkx genes are essential for maintenance of ventricular identity. Development 140, 4203–4213.

Tohyama, S., Hattori, F., Sano, M., Hishiki, T., Nagahata, Y., Matsuura, T., Hashimoto, H., Suzuki, T., Yamashita, H., Satoh, Y., et al. (2013). Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. Cell Stem Cell 12, 127–137.

Tomaru, Y., Hasegawa, R., Suzuki, T., Sato, T., Kubosaki, A., Suzuki, M., Kawaji, H., Forrest, A.R.R., Hayashizaki, Y., FANTOM Consortium, et al.

(2014). A transient disruption of fibroblastic transcriptional regulatory network facilitates trans-differentiation. Nucleic Acids Res *42*, 8905–8913.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

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

Wang, D., Passier, R., Liu, Z.P., Shin, C.H., Wang, Z., Li, S., Sutherland, L.B., Small, E., Krieg, P.A., and Olson, E.N. (2002). Regulation of cardiac growth and development by SRF and its cofactors. Cold Spring Harb. Symp. Quant. Biol. *67*, 97–105.

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

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

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics *16*, 284–287.

Zhou, H., Morales, M.G., Hashimoto, H., Dickson, M.E., Song, K., Ye, W., Kim, M.S., Niederstrasser, H., Wang, Z., Chen, B., et al. (2017). ZNF281 enhances cardiac reprogramming by modulating cardiac and inflammatory gene expression. Genes Dev. *31*, 1770–1783.

Zhu, W.Z., Van Biber, B., and Laflamme, M.A. (2011). Methods for the derivation and use of cardiomyocytes from human pluripotent stem cells. Methods Mol. Biol. *767*, 419–431.

Zhu, W.Z., Xie, Y., Moyes, K.W., Gold, J.D., Askari, B., and Laflamme, M.A. (2010). Neuregulin/ErbB signaling regulates cardiac subtype specification in differentiating human embryonic stem CellsNovelty. Circ. Res. *107*, 776–786.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Cardiac Troponin T antibody	Abcam	Cat# ab45932; RRID: AB_956386
Cardiac Troponin T Monoclonal Antibody (13-11)	Fisher	Cat# MA5-12960; RRID: AB_11000742
Goat anti-Rabbit IgG, Alexa fluor 594	Invitrogen	Cat# A11037; RRID: AB_2534095
Alexa Fluor® 488-conjugated AffiniPure Donkey Anti-Mouse IgG (H+L)	Jackson Immuno	Cat# 715-545-150; RRID: AB_2340846
TRA-1-60 Alexa Fluor™ 488	Invitrogen	Cat# A25618; RRID: AB_2885001
StainAlive TRA-1-60 Antibody (DyLight 488), Mouse anti-Human	Stemgent	Cat# 09-0068; RRID: AB_2233143
Bacterial and Virus Strains		
Adeno H5'.040.CMV.PI.Cre	Penn Vector Core	Adeno H5'.040.CMV.PI.Cre
Chemicals, Peptides, and Recombinant Proteins		
rhFGF-basic	Promega	Cat# G5071
Geltrex™ LDEV-Free, hESC-Qualified, Reduced Growth Factor Basement Membrane Matrix	Fisher	Cat# A1413301
Thiazovivin	Sigma	Cat# SML1045-5MG
Chiron 99021	Cayman Chemical	Cat# 13122
Activin A	R&D systems	Cat# 338-AC- 010
BMP4	PeproTech	Cat# AF-120-05ET
XAV 939	Tocris Bioscience	Cat# 3748
CHIR99021	Tocris	Cat# 4423
IWP 2	Tocris	Cat# 3533
Perturbation panel drugs	Selleck Chemicals	See Table S1
Critical Commercial Assays		
RNAtag-seq library prep protocol components	Shishkin et al., 2015	N/A
NEBNext Poly(A) mRNA Magnetic Isolation Module	NEB	Cat# E7490L
NEBNext Ultra II RNA Library Prep Kit for Illumina	NEB	Cat# E7770L
NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) oligos	NEB	Cat# E7600S
Illumina NextSeq 550 75 cycle high-output kit	Illumina	Cat# 20024906
Vector Red Substrate kit	Vector Labs	Cat# SK-5100
Deposited Data		
GTEx V7 RNA-seq TPMs	The GTEx Consortium, https://www.gtexportal.org/home/datasets	GTEx_Analysis_2016-01-15_v7_RNASeQCv1.1.8_gene_tpm.gct.gz
Tissue-specific regulatory networks FANTOM5-v1	Marbach et al., 2016	N/A
Perturbation panel profiling of human iPSC-derived cardiac myocytes and primary dermal fibroblasts	This paper	GSE167128
Gene expression after knockdown of transcription factors in human iPSC-derived cardiac myocytes	This paper	GSE166823
Experimental Models: Cell Lines		
Human: Platinum-A	Cell Biolabs	Cat# RV-102; RRID: CVCL_B489
Human: HEK293FT	Fisher	RRID: CVCL_6911
Human: GM00942	Coriell	Cat# GM00942; RRID: CVCL_9W78
Human: GM11169	Coriell	Cat# GM11169; RRID: CVCL_5P41
Human: immortalized cardiac fibroblasts ("immHCF")	Mohamed et al., 2017	N/A
Human: hiF-T iPSC-derived fibroblasts	Cacchiarelli et al., 2015	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
siRNAs targeting transcription factors and controls	See Table S2	N/A
single-molecule RNA FISH probe sets targeting NPPA, TNNT2, GAPDH	See Table S3	N/A
primers for amplification and cloning of transcription factors	See Table S3	N/A
RT-PCR primers for transcription factors	See Table S3	N/A
Recombinant DNA		
pLKO.1-based shRNAs targeting transcription factors and controls	See Table S2	N/A
pMXs-gw	Takahashi and Yamanaka, 2006	Addgene Plasmid #18656
pMXs-DsRed Express	Hong et al., 2009	Addgene Plasmid #22724
pMXs-GATA4, pMXs-MEF2C, pMXs-TBX5, pMXs-MESP1, pMXs-ESRRG, pMXs-MYOCD, pMXs-ZFPM2	Fu et al., 2013	N/A
pMD2.G	Trono lab (unpublished)	Addgene Plasmid #12259
psPAX2	Trono lab (unpublished)	Addgene Plasmid #12260
pMXs-SP3, pMXs-ZBTB10, pMXs-ZBTB44, pMXs-NFIA, pMXs-SSH2, pMXs-ZNF770, pMXs-ZFP91, pMXs-TurboGFP	This paper	N/A
Software and Algorithms		
STAR v2.5.2a	Dobin et al., 2013	N/A
HTSeq v0.6.1	Anders et al., 2015	N/A
DESeq2_1.24.0	Love et al., 2014	N/A
clusterProfiler_3.12.0	Yu et al., 2012	N/A
rajlabimagetools changeset 775fd10	Raj et al., 2008	N/A
All code used to produce figures in this paper	This paper	https://doi.org/10.5281/zenodo.5068731
Raw imaging and qPCR data	This paper	https://www.dropbox.com/sh/2ny7k6c4zy6zdsh/AADLNoom3YOw0Ps0B8w509R7a?dl=0

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Arjun Raj (arjunrajlab@gmail.com).

Materials availability
Materials generated in this study are available from the lead contact upon request. There are no restrictions to the availability of these materials.

Data and code availability
- All RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Microscopy data reported in this paper are publicly available via Dropbox as of the date of publication. Dropbox URL is listed in the key resources table. This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All cell culture incubations were performed at 37°C, 5% CO$_2$. We tested intermittently for mycoplasma contamination. Culture conditions for different experiments are specified in the method details section. In this study, we used the following human cell sources:

- GM00942 (human female skin fibroblasts; Coriell Cat# GM00942, RRID:CVCL_9W78)
- GM11169 (human cardiac fibroblasts; Coriell Cat# GM11169, RRID:CVCL_5P41)
- immHCF (immortalized human cardiac fibroblast line; Mohamed et al., 2017)
- hiF-T (human male iPSC-derived fibroblasts including doxycycline-inducible OSKM; Cacchiarelli et al., 2015)
- Platinum-A (transformed human female retroviral packaging cell line; RRID:CVCL_B489)
- HEK293FT (transformed human female lentiviral packaging cell line; RRID:CVCL_6911)

METHOD DETAILS

Cell culture
Unless otherwise noted, all cell culture incubations below were performed at 37°C, 5% CO$_2$. We tested intermittently for mycoplasma contamination.

GM00942 human dermal fibroblast culture
We cultured GM00942 human dermal fibroblasts (Coriell, GM00942; XX donor, normal-appearing tissue) according to the distributor's instructions, on tissue culture-treated dishes in E-MEM (QBI 112-018-101) + 10% FBS (Life Technologies 16000044, lot 1802004) + Pen/Strep.

GM11169 human cardiac fibroblast culture
We cultured GM11169 human cardiac fibroblasts (Coriell, GM11169; XX donor) on tissue culture-treated dishes in DMEM w/Glutamax + 9% FBS (Life Technologies 16000044) + P/S.

HEK293FT culture
We expanded HEK293FT cells in DMEM w/Glutamax + 9% FBS + P/S.

hiF-T culture
We cultured hiF-T cells as previously described prior to hiF-T-iPSC reprogramming experiments (Cacchiarelli et al., 2015). Briefly, we expanded hiF-T cells in growth medium on TC plastic dishes coated with Attachment Factor (Fisher S006100), and split cells 1:3 when they reached 60-70% confluency. hiF-T growth medium (GM) is DMEM/F-12 w/ Glutamax (Life Tech. 10565018) + 10% ES-FBS (Life Tech. 16141079) + 1x 2-Mercaptoethanol (Life Tech. 21985023) + 1x NEAA (Invitrogen 11140050) + P/S + 0.5μg/mL Puromycin + 16ng/mL rhFGF-basic (Promega G5071).

Immortalized human cardiac fibroblast (immHCF) culture
We cultured immortalized human cardiac fibroblasts (HCFs) as previously described (Mohamed et al., 2017). We received HCFs from Deepak Srivastava (Gladstone Institutes/UCSF). In brief, we expanded HCFs on gelatin-coated (Millipore ES-006-B) tissue culture dishes in iCM medium (per 500mL: 350mL DMEM w/Glutamax + 85mL Medium 199 + 50mL FBS + 5mL Non-essential amino acids + 10mL Pen/strep).

Platinum-A (Plat-A) retroviral packaging cell line culture
We cultured Platinum-A cells (Cell Biolabs RV-102) according to the manufacturer's instructions. We expanded these cells in DMEM w/Glutamax + 9% FBS + P/S + 10ug/mL Blasticidin + 1μg/mL Puromycin.

Cellular reprogramming
Derivation of cardiac myocytes (iPSC-CMs) from iPS-GM942-SeV3 iPSCs
We differentiated cardiomyocytes from iPS-GM942-SeV3 cells to create iPS-GM942-SeV3-CMs, as previously described (Laflamme et al., 2007; Palpant et al., 2015; Shanmughapriya et al., 2018; Zhu et al., 2010, 2011). Throughout this manuscript we refer to iPS-GM942-SeV3-CMs as 'iPSC-derived cardiac myocytes' or 'iPSC-CMs'. Briefly:
Seeding of iPS-GM942-SeV3
We thawed iPS-GM942-SeV3 and grew them in feeder-free conditions on Geltrex-coated (Thermo Fisher, cat. A1413301) dishes in StemMACS iPS Brew-XF (Miltenyi, cat. 130-104-368) for 4–5 days, until ~75% confluency. Then, we split and seeded iPS-GM942-SeV3 into Geltrex-coated 12-well plates at a density of 3 x 10^5 cells per well in iPS-Brew + 2μM Thiazovivin (Sigma, cat. SML1045-5MG). After 24 hours, we changed the culture medium to iPS-Brew + 1μM Chiron 99021 (Cayman Chemical, cat. 13122) and then incubated the cells for an additional 24 hr.

Differentiation to iPS-GM942-SeV3-CMs (a.k.a. iPSC-CMs)

Starting on Day 0, we incubated iPS-GM942-SeV3 for 18 hours in RPMI (Life Technologies, cat. 11875-119) + 100ng/ml Activin A (R&D systems, cat. 338-AC- 010) + 2% B-27 (minus insulin; Life Technologies, cat. 17504-044). Next, on Day 1 we changed the medium to RPMI + 2% B-27 (minus insulin) + 5ng/ml BMP4 (Peprotech, cat. AF-120-05ET) + 1uM Chiron 99021 and incubated these cells for 48 hours. On Day 3 we changed the medium to RPMI + 2% B-27 (minus insulin) + 1uM Xav 939 (Tocris Bioscience, cat. 3748) and incubated the cells for 48 hours. On Day 5 we changed the medium to RPMI + 2% B-27 (minus insulin) and incubated the cells for 72 hours. From Day 8 - 12 we changed the medium to RPMI + 2% B-27 (including insulin) + 1% pen/strep, and replaced the medium every other day.

Glucose-free medium selection steps for cardiac myocytes

In order to enrich the culture for cardiac myocytes, we subjected these cells to two low glucose selection steps. On Day 12 we started the first selection step by replacing the medium with RPMI glucose free (ThermoFisher, cat. 11879020) + 2% B-27 (including insulin) + 1% Pen-Strep and incubated for 72 hours. On Day 15 we replated the cells onto Geltrex-coated dishes at 6.3e5 cells/cm^2 in RPMI + 20% FBS (Seradigm, lot 050B14) + 1uM Thiazovivin and incubated them for 24 hours. On Day 16 we changed the medium to RPMI + 2% B-27 (including insulin) and incubated them for 48 hours to recover. On Day 18 we started the second selection step, again by replacing the medium with RPMI glucose free (ThermoFisher, cat. 11879020) + 2% B-27 (including insulin) + 1% Pen-Strep and incubated for 72 hours.

Replating iPS-GM942-SeV3-CMs into 96-well plates

On Day 21 of differentiation, we passaged the iPS-GM942-SeV3-CMs into Geltrex-coated 96-well plates at a density of 1e5 cells per well in RPMI + 20% FBS + 1μM thiazovivin and incubated overnight. On Day 22 we changed the medium to RPMI + 2% B-27 (with insulin) + antibiotics and cultured for an additional 48 hours to allow the plated cultures to start contracting again. By Day 23 or 24 we expected to observe recovery of contractile activity among a majority of the cardiac myocytes in all of the wells. If we did not observe beating activity we changed the RPMI + 2% B-27 (with insulin) + pen/strep medium on Day 24, incubated for another 48 hours, and checked again for contractile activity on Day 26. If the cardiac myocytes did not regain contractile activity by Day 26, we did not proceed to perturbation culture.

Perturbation culture
GM00942 fibroblasts

For GM00942 fibroblasts, we called the day on which they were seeded in 96-well plates Day -2 of perturbation. We split 95% confluent 10cm tissue culture-treated dishes of cells into 96-well plates at a density of roughly ~1.5e4 cells per well in EMEM + 10% FBS + Pen/strep and incubated for 48 hours. On Day 0 we replaced the medium (250 μL) in each well and added 0.7 μL of drug stock in DMSO (see Table S1) or of DMSO (for control cultures). This kept total DMSO concentration of the perturbation culture medium below 0.3% for all conditions. We incubated cells for 48 hours, and then on Day 2 replaced medium and re-adding a fresh dose of drug stock at the same volumes as Day 0. We incubated cells for a further 48 hours before taking images of each well (below) and extracting RNA (below) on Day 4. Per plate, all samples were located in annotated randomized well positions instead of grouping per-condition.

iPSC-derived cardiac myocytes

For iPSC-CMs, we called the first day on which we observed beating activity in the majority of wells of each 96-well plate Day 0 of perturbation. On Day 0 we replaced the medium with 250μL RPMI + 2% B-27 (with insulin) + pen/strep and added 0.7μL of drug stock in DMSO (See Table S1) or of DMSO (for control cultures), again keeping total DMSO concentration of perturbation culture medium below 0.3% for all conditions. We incubated cells for 48 hours, and then on Day 2 replaced medium and re-adding a fresh dose of drug stock at the same volumes as Day 0. We incubated cells for a further 48 hours before taking videos of each well (below) and extracting RNA (below) on Day 4. Per plate, all samples were located in annotated randomized well positions instead of grouping per-condition.

Transdifferentiation of fibroblasts to induced cardiac myocyte-like cells

We performed transdifferentiation of immortalized HCFs (immHCF), GM11169 fibroblasts, and GM00942 fibroblasts to induced cardiac myocyte-like cells as previously described(Mohamed et al., 2017). Briefly, on day -3 we plated 10^4 fibroblasts per well in 12-well culture vessels in iCM medium and Plat-A cells in 10cm dishes (4e6 cells per dish) in DMEM w/Glutamax + 9% FBS without any antibiotics. On day -2 we transfected each dish of Plat-A cells with 10ug of one indicated pMXs expression plasmid in 500uL Optimem + 35uL Fugene HD. On day 0, we collected viral supernatants and pooled them as needed for replicate conditions, filtered them through 0.45μm filter units, and transduced fibroblasts. For transductions we used 6μg/mL polybrene, a 30min 930 x g spin, and overnight incubation at 37C. On day 1 we replaced transduction medium with iCM medium. On day 4 we replaced iCM medium with 75% iCM medium/25% Reprogramming medium (RPMI 1640 + B-27 + P/S), on day 7 with 50% iCM medium/50% Reprogramming medium, on day 11 with 25% iCM medium/75% Reprogramming medium, and on day 14 with Reprogramming medium alone. We then changed reprogramming medium daily until analysis on the indicated day per experiment, usually day 24. On day 24, we fixed cells in two formats for analysis. Some 12-well TC plastic wells were fixed in place using 3.7% formaldehyde and permeabilized at least overnight with 70% Ethanol in 4C, while others were dissociated with Accutase and transferred to Concanavalin-coated (Sigma C0412) 8-well Lab-Tek chambers. After 90-120 minutes, transferred samples were fixed in these 8-well chambers using 3.7% formaldehyde and permeabilized at least overnight with 70% ethanol in 4C. Samples in 12-well wells were processed for FISH imaging by

excising them from their 12-well plate after fixation with a heated 20mm cork borer and processing them for FISH or immunofluorescence as described below.

Cloning of transcription factor genes and TurboGFP into pMXs

In order to drive overexpression of perturbable transcription factor genes and TurboGFP, we cloned cDNA for genes of interest into pMXs-gw (Addgene 18656; a gift from Shinya Yamanaka) using BP and LR Clonase II (Invitrogen). We amplified cDNA of targets of interest using attB-target-specific primers (Table S3). We used standard tools to verify sequence identity of the plasmid backbone and gene insert, such as restriction digestion and Sanger sequencing. We amplified attB-TurboGFP off of the SHC003 plasmid (Sigma SHC003).

Titering of pMXs retroviral vectors

Since our expression vectors do not contain selectable or fluorescent markers and pMXs retroviral vectors only transduce dividing cells, we indirectly titered each experimental replicate's batch of virus by co-transducing parallel samples of HCFs with pMXs-DsRed Express (Addgene 22724; a gift from Shinya Yamanaka) and pMXs-TurboGFP (see "Cloning" above) produced using the same batch of Plat-A cells under the same conditions. In order to estimate the fraction of cells that are infected at least once per transcription factor, we considered the infection rates of these fluorescent pMXs vectors. By comparing the fraction that are co-infected with both against the fraction that are infected with each transcription factor at all, we can infer the fraction of cells dividing in the population during the transduction period and the fraction of those cells that receive at least one copy of any individual expression vector. We make the simplifying assumption that among dividing cells infection events are independent of each other. Therefore, the ratio of the fraction that are DsRed+ and GFP+ to the fraction that are DsRed+ (or GFP+) is approximately the square of the transduction rate for any individual virus. E.g., for 30% of cells being DsRed+ and 24.3% being DsRed+ and GFP+, 24.3/30 = 0.81, which gives 90% transduction rate for each individual virus. We used two-color flow cytometry (GUAVA) to assess DsRed+, GFP+, and DsRed+ GFP+ fractions per sample.

Gene knockdown in iPSC-derived cardiomyocytes

Cardiac myocytes were obtained by differentiating induced pluripotent stem cells (iPSC; WTC-11 cell line) as described previously (Judge et al., 2017; Miyaoka et al., 2014). Briefly, undifferentiated iPS cells were maintained and seeded on mTesr media (StemCell Technologies). Three days later media was changed to RPMI with B-27 supplement (Gibco) without insulin supplemented with CHIR99021 (Tocris). After 48 hours, media was changed to RPMI with B-27 (minus insulin) supplemented with IWP2 (Tocris), and to RPMI with B-27 (with insulin) two days after that. Cultures displayed beating in at least 50% of the cell area by day 10. Cultures were then cardiomyocyte enriched by following a metabolic switch protocol, to an efficiency of >90% ACTN2 positive cells (Tohyama et al., 2013). Briefly, day 15 differentiation cultures were replated and treated with glucose-free media supplemented with GlutaMax (Gibco), Non Essential Amino Acids (Gibco) and 4mM of lactate (Sigma-Aldrich). Media was changed every other day for a total of 6 days. At day 30, cardiac myocytes were replated into a 24 well plate (100,000 cells/well) and allowed to recover for four days before knockdown.

   Oligos for siRNA knockdown were ordered from Horizon Discovery. A total of three oligos per target were pooled together. All knockdown experiments were performed in triplicate. Knockdown (day 0) was performed using the Lipofectamine RNAiMAX reagent (ThermoFisher) and pooled siRNA at a final concentration of 10uM. Media was changed to fresh media 48 hours later. Five days post knockdown, a second identical dose of lipofectamine+siRNA was added to the cells for another 48 hours. At day 10, videos of beating cardiac myocytes were automatically acquired using the Cellogy Pulse system (Dana Solutions) (Maddah et al., 2015). Then either we harvested RNA from cells using RLT buffer (Qiagen) supplemented with 1:100 beta-mercaptoethanol or we fixed the cells for immunostaining using 4% paraformaldehyde.

shRNA-mediated knockdown of transcription factors in hiF-T cells

We conducted knockdown of individual transcription factors using shRNAs essentially as previously described.(Cacchiarelli et al., 2015) In brief, we acquired cloned pLKO.1, pLKO.1-shRNA, and pLKO.1-TurboGFP plasmids from the University of Pennsylvania High-Throughput Screening Core (Table S2). We verified shRNA and backbone sequence with Sanger sequencing. We packaged shRNA lentivirus using pMD2.G (Addgene 12259; a gift from Didier Trono) and psPAX2 (Addgene 12260; a gift from Didier Trono) in HEK293FT cells, and filtered viral supernatant through 0.22μm filter units prior to infecting hiF-T cells. We infected hiF-T cells at an MOI of approximately 1 (for a transduction efficiency of ~70%) with 4μg/mL polybrene and 30 min 930 x g centrifugation. Since hiF-T cells are already Puromycin-resistant, we were unable to perform an antibiotic selection step after infection with these pLKO.1-puro-based shRNA plasmids.

Verification of knockdown efficiency following shRNA transduction

We performed RT-qPCR on RNA extracted from samples of the hiF-T cells that we used in reprogramming experiments. We used Superscript III Reverse Transcriptase for first-strand cDNA synthesis and Power SYBR qPCR Master Mix with gene-specific primer pairs for qPCR on an Applied Biosystems 7300 system. We performed all statistical analysis using custom scripts in R (see "Code accessibility" below for all scripts) and calculated knockdown efficiency using the ΔΔCt method.

hiF-T reprogramming to pluripotency

We performed hiF-T reprogramming experiments as previously described. Briefly, after shRNA transduction on day -7, we expanded cells in hiF-T GM without puromycin for one week. On Day -1 we seeded CF-1 Irradiated MEFs on uncoated 24-well plates (Corning) at a density of 2.5 x 10^5 cells per well in hiF-T GM without puro. On Day 0, we seeded 10^4 hiF-T cells per 24-well plate well. On Day 1 we began Yamanaka factor induction by switching media to hiF-T GM with 2μg/mL doxycycline and without puromycin. On Day 3 we switched media to KSR Medium (KSRM): DMEM/F-12 w/ Glutamax (Life Tech. 10565018) + 20% Knockout Serum Replacement (Life Tech. 10828010) + 1x 2-Mercaptoethanol (Life Tech. 21985023) + 1x NEAA (Invitrogen 11140050) + P/S + 8ng/mL rhFGF-basic + 2μg/mL Doxycyclin. We changed KSRM daily, and analyzed cells on day 21. We performed a total of 3 biological replicates (i.e., different vials of hiF-T cells expanded and reprogrammed on different days with different batches of media), each with technical triplicates per shRNA condition. All biological and technical replicates of all reprogramming experiments we conducted are presented in this manuscript.

High-throughput RNA extraction

We used RNaqueous-96 kits (Ambion AM1920) for RNA extraction without the optional DNase step, according to manufacturer's instructions.

RNAtag sequencing

We conducted highly parallelized bulk RNA sequencing with RNAtag-seq as previously described, using all components and steps in the published protocol (Shishkin et al., 2015) for all perturbation culture samples. We ordered the specified 32 barcoded DNA oligos for RNAtags from Biosearch Technologies and indexed primers for library amplification and reverse transcription from IDT. We sequenced all RNAtag-seq libraries in batches of 96 samples on an Illumina NextSeq 550 using 75 cycle high-output kits (Illumina 20024906).

RNA sequencing

We conducted standard bulk paired end (37:8:8:38) RNA sequencing as previously described for iPSC-CM knockdown samples using RNeasy Mini (Qiagen 74104) for RNA extraction, NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB E7490L), NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB E7770L), NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) oligos (NEB E7600S), and an Illumina NextSeq 550 75 cycle high-output kit (Illumina 20024906).

RNAtag-seq and RNA-seq data processing

We demultiplexed RNAtag-seq reads using custom scripts, courtesy of Edward Wallace (https://github.com/ewallace/pyRNATagSeq, changeset 6ffd465). Then as previously described we aligned RNAtag-seq reads to the human genome (hg19) with STAR v2.5.2a and counted uniquely mapping reads with HTSeq v0.6.1 (Anders et al., 2015; Dobin et al., 2013; Shaffer et al., 2017).

Live cell Tra-1-60 imaging

In a pilot reprogramming experiment without shRNA transduction, we conducted live-cell staining of hiF-T-iPSC colonies Tra-1-60 with TRA-1-60 Alexa Fluor™ 488 Conjugate Kit for Live Cell Imaging (Life Tech. A25618) according to the manufacturer's instructions.

Alkaline phosphatase staining with colorimetry

We used the Vector Red Substrate kit (Vector Labs SK-5100) to stain hiF-T-iPSC colonies after fixation on day 21 of reprogramming experiments. We fixed wells in 24-well format using 3.7% formaldehyde for 3 minutes, and followed the manufacturer's instructions.

Immunofluorescence

We performed immunofluorescence for several markers. For Tra-1-60 immunofluorescence of hiF-T-iPSC samples that had already been stained with Vector Red, we blocked and permeabilized in 5% BSA + 0.1% Triton X-100 in PBS at room temperature for 30 min. Then we washed samples in PBS and used Stemgent 09-0068 at 1:200 in 5% BSA + 0.1% Triton X-100 for 2 hours at room temp. We washed samples in PBS and stained with DAPI prior to imaging. For cardiac troponin immunofluorescence of iPSC-CM and trans-differentiated samples, we fixed samples in 3.7% formaldehyde for 10 min at room temp, washed in PBS, and permeabilized with 70% ethanol overnight at 4C. Independent of smFISH or after the smFISH protocol, we performed immunofluorescence with Abcam ab45932 primary (1:200) with goat anti-rabbit-Alexa 594 (1:200) secondary and Fisher MA5-12960 primary (1:200) with donkey anti-mouse-Alexa 488 (1:200) secondary. We used samples in 3% BSA + 0.1% Tween 20 for blocking/binding buffer. Primary antibody incubations of 1 hour and secondary incubations of 30 min, both at room temperature. Samples were washed with PBS and stained with DAPI prior to imaging.

Single-molecule RNA FISH

GAPDH probes were used as previously described (Padovan-Merhar et al., 2015). We incubated our cells overnight at 37°C in hybridization buffer (10% dextran sulfate, 2× SSC, 10% formamide) with standard concentrations of RNA FISH probes (Table S3) (Raj et al., 2008). The following morning, we performed two washes in wash buffer (2X SSC, 10% formamide), each consisting of a 30-min

incubation at 37°C. After the second wash, we rinsed once with 2X SCC/DAPI and mounted the sample for imaging in and 2X SSC (Raj et al., 2008). We performed RNA FISH on cell culture samples grown on a Lab-Tek chambered coverglass using 50 μL of hybridization solution spread into a thin layer with a coverslip and placed in a parafilm-covered culture dish with a moistened Kimwipe to prevent excessive evaporation.

Imaging
We imaged each sample on a Nikon Ti-E inverted fluorescence microscope using a 60X Plan-Apo objective and a Hamamatsu ORCA Flash 4.0 camera. For 60X imaging of complete cells, we acquired z-stacks (0.3 μm spacing between slices). For 60X imaging of a large field of cells with one plane each, we used Nikon Elements tiled image acquisition with perfect focus. All images of stained cells were in different fluorescence channels using filter sets for DAPI, Atto 488, Cy3, Alexa 594, and Atto 647N. The filter sets we used were 31000v2 (Chroma), 41028 (Chroma), SP102v1 (Chroma),17 SP104v2 (Chroma) and SP105 (Chroma) for DAPI, Atto 488, Cy3, Atto 647N/Cy5 and Atto 700, respectively. A custom filter set was used for Alexa 594/CalFluor610 (Omega). We tuned the exposure times depending on the dyes used: 400 ms for probes in Cy3 and Alexa 594, 500 ms seconds for each probe in Atto 647N, and 50 ms for DAPI probes. We also acquired images in the Atto 488 channel with a 400 ms exposure as a marker of autofluorescence.

Image Processing
smFISH analysis of image scans and stacks was done as previously described using rajlabimagetools changeset 775fd10 (https://bitbucket.org/arjunrajlaboratory/rajlabimagetools/wiki/Home) in MATLAB v2019a, and is compatible with rajlabimagetools changeset a2c6ac5 (https://github.com/arjunrajlaboratory/rajlabimagetools/) in MATLAB v2017a (Raj et al., 2008).

Reproducible analysis
For all raw imaging and qPCR data, please see: https://www.dropbox.com/sh/2ny7k6c4zy6zdsh/AADLNoom3YOw0Ps0B8w509R7a?dl=0. For all raw sequencing data, please see GEO accessions GSE167128 and GSE166823. For a fully reproducible processing pipeline, including raw qPCR data, from all raw data types to the graphs and images included in this paper, please see: https://doi.org/10.5281/zenodo.5068731. For details, please see the readme files included in each folder.

QUANTIFICATION AND STATISTICAL ANALYSIS

We performed analyses of RNAtag-seq counts, RNA-seq counts, fibroblast-to-iPSC reprogramming results, and fibroblast-to-cardiac myocyte transdifferentiation results in R v3.6.1 using packages e1071_1.7-3, gridExtra_2.3, DESeq2_1.24.0, SummarizedExperiment_1.14.1, DelayedArray_0.10.0, BiocParallel_1.18.1, matrixStats_0.57.0, GenomicRanges_1.36.1, GenomeInfoDb_1.20.0, metaRNASeq_1.0.3, org.Hs.eg.db_3.8.2, AnnotationDbi_1.46.1, IRanges_2.18.3, S4Vectors_0.22.1, Biobase_2.44.0, BiocGenerics_0.30.0, clusterProfiler_3.12.0, readxl_1.3.1, ggrepel_0.8.2, magrittr_1.5, forcats_0.5.0, stringr_1.4.0, dplyr_1.0.2, purrr_0.3.4, readr_1.4.0, tidyr_1.1.2, tibble_3.0.3, ggplot2_3.3.2, tidyverse_1.3.0, yaml_2.2.1, and their associated dependencies (Love et al., 2014; Yu et al., 2012).

Details of particular statistical tests, definitions of significance, and the number and type of replicates for all analyses can be found in the corresponding figure legend. Unless otherwise noted, error bars represent standard error of the mean.

Gene expression perturbation-responsiveness
As a measure of gene expression perturbation-responsiveness we used the count of the number of conditions in which a gene was differentially expressed relative to cell type DMSO controls. For most analyses we used any change with a DESeq2 adjusted p-value less than 0.1, but also conducted analyses with additional filters, such as minimum absolute values of log2FoldChange. Filter criteria are explicitly noted in figure legends for each analysis. For responsiveness analyses we used a minimum average RPM of 20 in control samples based on sequencing depth analyses in Figure S4, with additional filters on TPM used as marked in particular analyses. We also considered other measures of perturbation-responsiveness, as well, which are not explored in the manuscript above, details of which are available upon request.

Gene expression cell type-specificity analysis
We calculated Jensen-Shannon divergence-based scores for each gene as previously described (Cabili et al., 2011) in each tissue or cell type in the GTEx V7 RNA-seq dataset (GTEx Consortium et al., 2017). Briefly, we analyzed all V7 GTEx RNA-seq samples with: RIN (RNA integrity score) $\geq$ 7, mapping rate $\geq$ 50%, and 1 million or more mapped reads. With these filters, 3623 samples remained in the processed GTEx dataset. We then calculated JS_{sp} values for each gene in each tissue type (SMTS identifier) based on average expression levels per tissue type in transcripts per million (TPM), as described in Cabili et al. (2011). For the analysis presented in main text figures, we removed GTEx Skin and Heart from the dataset and added GM00942 fibroblast and iPSC-CM cardiac myocyte control average expression expression values. We present specificity score calculations with and without removal of GTEx Skin and Heart and with and without addition of GM00942 fibroblast and iPSC-CM cardiac myocyte data in the supplement.

Gene set enrichment analysis
For perturbation culture samples, we performed GO-Cellular Component gene set over-enrichment analysis on highly expressed genes (mean RPM > 20) for each cell type using the clusterProfiler v3.12.0 package in R v3.6.1 (Yu et al., 2012). We compared

iPSC-CM-responsive highly expressed genes (up-regulated in 4 or more conditions and mean RPM > 20 in control iPSC-CM) against the entire set of genes expressed > 20 RPM in iPSC-CM controls. We compared iPSC-CM-unresponsive genes (up-regulated in 0 or 1 conditions and mean RPM > 20 in control iPSC-CM) against the entire set of genes expressed > 20 RPM in iPSC-CM controls. We compared fibroblast-responsive highly expressed genes (up-regulated in 4 or more conditions and mean RPM > 20 in control GM00942 fibroblasts) against the entire set of genes expressed > 20 RPM in GM00942 controls. We compared fibroblast-unresponsive genes (up-regulated in 0 or 1 conditions and mean RPM > 20 in control GM00942) against the entire set of genes expressed > 20 RPM in GM00942 controls.

For iPSC-CM knockdown samples, we performed GO-Cellular Component gene set enrichment analysis on up- or down-regulated genes in each gene-targeting siRNA condition compared against scrambled control samples. We filtered GO terms to level 5 and used the clusterProfiler::simplify() function to consolidate reported terms.

Meta-analysis of differential expression
We used Fisher's method with Benjamini-Hochberg adjustment, with FDR = 0.05, to combine adjusted p-values from DESeq2 for each gene across the perturbation panel in iPSC-CMs. This tested the null hypothesis that there was no differential expression caused by any perturbation for each gene.

Transcription factor regulon responsiveness analysis
We analyzed transcription factor target set (i.e., regulon) responsiveness in iPSC-CM and fibroblast perturbation culture samples for transcription factors and their predicted targets expressed > 20 RPM in iPSC-CM and fibroblast controls, respectively. We used published, annotated regulons from a recent study that inferred direct regulatory relationships in each human tissue for 662 transcription factors using a combination of the FANTOM5 cap analysis of gene expression dataset and a curated set of transcription factor sequence binding motifs (Marbach et al., 2016). We used the "high-level" "heart" gene regulatory network for regulon responsiveness analysis in iPSC-CMs and the "high-level" "connective tissue integumental cells" network and the "individual-level" "fibroblast-dermal" and "fibroblast-skin normal" networks for analysis in fibroblasts. From these network, for each transcription factor expressed > 20 RPM in the respective control samples in this study, we extracted edge-weights for each predicted target-regulatory interaction. Then, we calculated an overall regulon responsiveness score for each expressed transcription factor in the network by summing over the products of each predicted target's number of conditions in which it was up-regulated and the fraction of differential expression conditions in which it was up-regulated, normalized by the edge-weights.

Prioritization of highly responsive genes for use in reprogramming experiments
For iPSC reprogramming experiments we considered transcription factor genes (Vaquerizas et al., 2009) that were 1) up-regulated in at least 5 conditions in GM00942, 2) up-regulated in at least 50% more conditions than they were down-regulated in GM00942, 3) not also frequently up-regulated in iPSC-CM, 5) expressed at >50 TPM in GM00942 controls as a filter for the highest expressed transcription factors, 6) not commonly studied in the context of fibroblast development based on literature review, 7) not commonly considered a member of a stress response or apoptosis pathway based on literature review using Google Scholar search terms including "fibroblast differentiation", "skin development", "stress", and the Gene Symbols of interest, and 8) with at least 3 quality-controlled targeting shRNA clones available through our university core service lab's Human TRC 2.0 lentivirus library. Ultimately we tested knockdown of the genes 'SKIL', 'YBX3', 'TSC22D1', 'CERS2', 'KLF13', 'TBX3', 'ID1', 'ATOH8', 'ZNF652', 'NFATC4', 'ZBTB38', 'LARP1', 'CEBPB', 'ID3', 'PRRX2', and 'RUNX1'.

For cardiac transdifferentiation experiments we considered transcription factor genes (Vaquerizas et al., 2009) that were 1) up-regulated in at least 4 conditions in iPSC-CM, 2) up-regulated in at least 50% more conditions than they were down-regulated in iPSC-CM, 3) not also frequently up-regulated in GM00942, 5) expressed at >50 TPM in iPSC-CM controls as a filter for the highest expressed transcription factors, 6) not commonly studied in the context of fibroblast development based on literature review, and 7) not commonly considered a member of a stress response or apoptosis pathway based on literature review. Ultimately we tested overexpression of the genes 'SP3', 'ZBTB10', 'ZBTB44', 'SSH2', 'NFIA', 'ZNF652', and 'ZFP91'.