Trajectory reconstruction identifies dysregulation of perinatal maturation programs in pluripotent stem cell-derived cardiomyocytes

Suraj Kannan¹,²,³, Matthew Miyamoto¹,²,³, Renjun Zhu¹,²,³, Michaela Lynott⁴, Jason Guo¹,²,³, Elaine Zhelan Chen¹,²,³, Alexandre R. Colas⁴, Brian Leei Lin¹, Chulan Kwon¹,²,³,⁵,*
¹Division of Cardiology, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA
²Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD, USA
³Institute for Cell Engineering, Johns Hopkins School of Medicine, Baltimore, MD, USA
⁴Sanford Burham Prebys Medical Discovery Institute, San Diego, CA, USA
⁵Lead contact

SUMMARY

A limitation in the application of pluripotent stem cell-derived cardiomyocytes (PSC-CMs) is the failure of these cells to achieve full functional maturity. The mechanisms by which directed differentiation differs from endogenous development, leading to consequent PSC-CM maturation arrest, remain unclear. Here, we generate a single-cell RNA sequencing (scRNA-seq) reference of mouse in vivo CM maturation with extensive sampling of previously difficult-to-isolate perinatal time periods. We subsequently generate isogenic embryonic stem cells to create an in vitro scRNA-seq reference of PSC-CM-directed differentiation. Through trajectory reconstruction, we identify an endogenous perinatal maturation program that is poorly recapitulated in vitro. By comparison with published human datasets, we identify a network of nine transcription factors (TFs) whose targets are consistently dysregulated in PSC-CMs across species. Notably, these TFs are only partially activated in common ex vivo approaches to engineer PSC-CM maturation. Our study can be leveraged toward improving the clinical viability of PSC-CMs.

Graphical abstract

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Correspondence: ckwon13@jhmi.edu.

AUTHOR CONTRIBUTIONS
S.K., B.L.L., and M.M. were directly responsible for performing experiments in this manuscript. All authors provided significant intellectual contribution to the manuscript. The manuscript was initially written by S.K. with subsequent revisions and input from all authors. All authors approve of the final manuscript.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112330.

DECLARATION OF INTERESTS
The authors declare no competing interests.
In brief

By directly comparing single-cell RNA sequencing trajectories of in vivo and in vitro cardiomyocyte maturation, Kannan et al. find that pluripotent stem cell-derived cardiomyocytes fail to undergo a critical perinatal program, leading to maturation arrest.

INTRODUCTION

Pluripotent stem cell-derived cardiomyocytes (PSC-CMs) offer a powerful solution to numerous challenges in clinical cardiology, with applications including regenerative medicine, drug screening, and disease modeling. However, the inability of PSC-CMs to mature to an adult-like phenotype has precluded their effective biomedical use. A number of transcriptomic, proteomic, structural, and functional measurements have indicated that PSC-CMs more closely resemble fetal or embryonic CMs rather than their mature adult counterparts. While this phenomenon is often termed a “maturation arrest,” there is little mechanistic understanding of why PSC-CMs fail to recapitulate the adult phenotype. It is known that differentiating PSC-CMs faithfully undergo cascading gene expression changes associated with gastrulation, mesoderm induction, and cardiomyogenesis. However, it is unclear to what degree they initiate a maturation-like program in vitro or when and how this program is disrupted. This is further complicated by the lack of knowledge of the initiation, regulation, and dynamics of CM maturation in vivo. To address this, our previous work aimed to identify a transcriptional landscape for CM maturation over in vivo development.
We subsequently compared the establishment of gene regulatory networks (GRNs) between PSC-CMs and endogenous CMs. However, this work was done using bulk samples, while CM maturation occurs heterogeneously across time both in vivo and in vitro, suggesting a need for further data.

To address the limited maturation of PSC-CMs, a number of groups have developed ex vivo perturbation protocols to improve PSC-CM maturation. These approaches have included cytokine, growth factor, and hormone cocktails; co-culture with other cells; induction of physical stimuli (e.g., mechanical stretch, electrical stimulation); and construction of biomaterial-based three-dimensional tissues. Ostensibly, the goal of these perturbations is to replicate fundamental aspects of the native cardiac milieu to engineer maturation. However, assessment of the maturation of these perturbed tissues is often done through ad hoc phenotypic measurements. Additionally, there has been no systematic comparison of whether these perturbations activate maturation pathways analogous to endogenous development. Thus, it is unclear to what degree any individual perturbation is truly biomimetic. Other groups have shown that, despite their immature phenotype, PSC-CMs may already be capable of modeling specific (albeit limited) aspects of cardiac biology. This approach has led to breakthrough applications in drug screening and disease modeling. A deeper biological understanding of CM maturation processes, however, can significantly expand the possibilities for future clinical application of PSC-CMs.

In this study, we used single-cell RNA sequencing (scRNA-seq) to directly compare maturation processes between endogenous and PSC-CMs. Through use of our previously established large-particle fluorescence-activated cell sorting (LP-FACS) protocol, we established a high-quality reference of mouse CM maturation, in particular sampling previously understudied perinatal time points. We subsequently generated embryonic stem cell (ESC) lines from the same strain used for the in vivo reference and differentiated these to PSC-CMs to produce an in vitro maturation reference. We found that endogenous CMs undergo a perinatal maturation program between postnatal day (P)8–P15 that is poorly recapitulated in vitro. By cross-referencing with published human PSC-CM datasets, we identified a network of nine transcription factors (TFs) whose targets are consistently dysregulated in PSC-CMs across species. Through further meta-analysis of published perturbation RNA-seq datasets, we found that ex vivo perturbations can partially activate some of the key maturation-related TFs, but no method completely activates them all. Our study is the first to provide systematic single-cell comparison of maturation programs between in vivo and in vitro CMs and may open future avenues for generating fully mature PSC-CMs.

RESULTS

Trajectory reconstruction identifies perinatal window in endogenous CM maturation

Our first goal was to develop an improved understanding of gene expression changes in CM maturation during in vivo development. scRNA-seq of CM maturation, in particular postnatal maturation, has been previously limited due to difficulty in isolating large, fragile CMs. Recently, however, we established LP-FACS as a viable approach to generate high-
quality CM scRNA-seq libraries and used this technique to produce a reference dataset of CM maturation. In this study, we expanded our reference to a total of ~1,600 left ventricular free wall CMs encompassing 15 time points ranging from embryonic day (E)14 to P84 (Note S1). We specifically isolated CMs through use of Myh6-Cre; mTmG (aMHC x mTmG) mice, in which cells expressing cardiac-specific myosin heavy chain are readily separated by GFP expression (Figure 1A). We further validated CM identity of sequenced libraries through use of the SingleCellNet algorithm, which computationally classifies single cells by gene expression (Figure S1A). CMs were sequenced to a median depth of ~21,000 unique molecular identifiers (UMIs) per cell and ~4,000 genes per cell. This high-depth sequencing is particularly important for analyzing CMs undergoing maturation because of their gradual transcriptomic changes. Our dataset was designed to encompass the full temporal range of CM maturation while particularly sampling perinatal time points that may be critical to the maturation process.

To understand the gene expression changes over CM maturation at the single-cell level, we performed trajectory reconstruction using Monocle 3 (Figure 1B; Note S2). Monocle 3 recovered a unidirectional, non-branching trajectory, matching earlier reports. Trajectory reconstruction enables calculation of “pseudotime,” a metric of progression along an inferred biological process. We plotted the Monocle 3-computed pseudotimes at each biological time point, which validated that pseudotime progressively increased over biological time (Figure 1C). Notably, however, at each biological time point, there was significant heterogeneity of pseudotime, indicating that maturation proceeds asynchronously at single-cell level. Additionally, we observed several interesting transition points. From E14 to E18, there was a jump in pseudotime, marking the initiation of maturation. Between E18 and P4 (corresponding to pseudotime \( \in [5, 20] \)), individual cells proceeded through a late embryonic/neonatal phase of maturation, but the distribution of pseudotimes was similar at each biological time point. Subsequently, there was a large jump in pseudotime between P8 and P14 (corresponding to pseudotime \( \in (20, 40] \)), suggesting a perinatal maturation process. Finally, after P15 (corresponding to pseudotime \( \in (40, 60] \)), cells converged to a relatively mature phenotype, though there was still notable heterogeneity even in these late stages. Based on these results, we labeled the pseudotime intervals \( [0, 20] \), \( (20, 40] \), and \( (40, 60] \) as embryonic/neonatal, perinatal, and adult, respectively.

Using Monocle 3, we identified 3,015 genes with differential expression over pseudotime. Over 80% of these genes were downregulated over pseudotime, which supports our previous assertion that CM maturation involves significant pruning of unneeded gene modules. As expected, upregulated genes corresponded to crucial aspects of mature CM biology, including sarcomeric, calcium handling, and fatty acid metabolism, while downregulated genes were enriched for cell-cycle and transcription/translation processes (Figure S1B). We next aimed to identify when maturation genes become differentially expressed along pseudotime. To do this, we produced sequential subsets of our inferred trajectory by binning cells with pseudotime \( \in [0, 1] \), pseudotime \( \in [0, 2] \), etc., up to a bin of cells with pseudotime \( \in [0, 60] \) (e.g., the entire trajectory). We then computed differential genes for each subset to identify what percentage of the total 3,015 differentially expressed genes become differentially expressed within each progressive subset. 22% of genes become differentially expressed by pseudotime 20, while 80% become differentially expressed by
pseudotime 40 (Figure 1D). Thus, 58% of genes become differentially expressed within the perinatal period, further supporting this period as critical for CM maturation.

As a complementary approach to trajectory reconstruction, we investigated transcriptional dynamics by computing RNA velocity using the scVelo package. RNA velocity uses the ratio of spliced and unspliced messenger RNA reads in scRNA-seq data to determine the rate of gene expression changes. These computed velocities can be aggregated and projected onto developmental trajectories to study cell differentiation dynamics. Here, we computed velocities for our identified differentially expressed genes, and embedded these velocities on our Monocle 3-inferred trajectory (Figure 1E). The RNA velocity field pointed along the maturation trajectory in the embryonic and perinatal phases before subsequently becoming more incoherent. The loss of coherence of the velocity field in later time points likely indicates that these time points at dynamic equilibrium, with no clear transcriptional directionality. These results are further supported by the ratios of spliced to unspliced RNAs, which steadily decrease over biological time until becoming stable around P15 (Figure S1C). The rate of differentiation can be quantified by the length of the velocity vector. We found that the velocity length progressively increased starting at pseudotime 0, reaching a maximum for pseudotime $\in [20, 40]$, before subsequently decreasing (Figure 1F). These results indicate that, endogenously, CM maturation proceeds with highest velocity in the perinatal phase.

Taken together, our results support the existence of a perinatal window for CM maturation. Based on this, we limited our subsequent analysis to the perinatal period by selecting cells with pseudotime $\in [0, 42]$. We selected 42 as our upper pseudotime threshold to effectively capture our identified perinatal window ([20, 40]) plus a small margin of error. During this period, 2,628 genes become differentially expressed, which we subsequently refer to as maturation genes.

**Maturation genes form early-activating and late-activating clusters with unique upstream regulation**

We next sought to determine and isolate groups of genes that share temporal expression patterns from embryonic to perinatal stages. Though some previous methods have been developed for quantifying gene dynamics along pseudotime trajectories, we instead developed a flexible method using built-in features of Monocle 3. For each gene, Monocle 3 fits a smooth spline curve (Figure 1G). We defined the maturation fold change (FC) as the maximum FC that happens over this spline curve in our predetermined pseudotime interval of [0, 42]. We then identified the time to x% FC as the pseudotime at which x% of the maturation FC is achieved. The x% used can be calibrated to study different aspects of individual gene dynamics. For each maturation gene, we computed the time to 10% FC (gene activation/inactivation), time to 50% FC (midpoint of gene activity), and time to 95% FC (gene plateau).

Based on these parameters, we identified 5 gene clusters (Figures 1H–1J and S2; supplemental methods). Of these clusters, three showed time to 10% FC below pseudotime 5. Thus, we termed these clusters early-activating clusters 1, 2, and 3 (EA1, EA2, and EA3, respectively). We termed the remaining two clusters late-activating clusters 1 and 2 (LA1
and LA2, respectively). Our computed gene dynamics allowed for a better understanding of CM maturation-related gene changes. Given the prior evidence suggesting the importance of the perinatal window, we expected that a large number of maturation genes would initiate their changes at around pseudotime 20. By contrast, 78% of maturation genes fell into the early-activating categories, and only LA2 (3.4% of maturation genes) showed a time to 10% FC near the onset of the perinatal window. However, apart from EA1, all of the clusters showed time to 95% FC in the pseudotime range [30, 40]. In summary, maturation-related gene changes as a whole likely initiate early. The embryonic/neonatal period is characterized by completion of the EA1 gene program, while the perinatal period is characterized by completion of the EA2, EA3, and LA1 gene programs, as well as initiation and completion of the LA2 gene program. Notably, these dynamics are qualitatively different than those seen in cellular differentiation, which are often driven by “switch-like” genes.

Given the differences in dynamics for each of the gene clusters, we were curious to know whether different clusters had different upstream transcriptional regulators. For each cluster, we performed over-representation analysis of TF targets, and further investigated TFs with high fold enrichments across the clusters (Figure 1K). This analysis identified several TFs that have been previously implicated in CM maturation, including Srf, Err1, Arnt, and Mef2, as well as others with previously undescribed function. Interestingly, while some TFs showed stage-specific enrichment, a large number showed enrichment across multiple clusters, despite different transcriptional dynamics. In particular, inferred regulators of EA1 as well as the LA clusters often showed enrichment in other clusters (typically EA2 and EA3). Our results provide a first description of the timing and regulation of CM maturation in vivo.

**scRNA-seq recovers maturation trajectory in PSC-CMs**

Our next goal was to establish a complementary reference of PSC-CM maturation. In order to directly compare PSC-CMs with our in vivo CM trajectory, we generated four separate, isogenic ESC lines from aMHC x mTmG mice. We subsequently differentiated these to PSC-CMs using a protocol adapted from previous studies, with sequential Wnt modulation through defined small molecules (Figure 2A). Our rationale was that by using the same parent mouse line, our comparisons could minimize effects of confounding caused by strain/line differences and isolate biological differences between the in vivo and in vitro environments.

Differentiation of the ESC lines yielded beating GFP+ PSC-CMs by day 6.5–7 of differentiation, depending on the line. PSC-CMs additionally displayed other CM markers such as Tnnt2 and displayed sarcomeric structures (Figure 2B). We isolated PSC-CMs by conventional FACS from 8 time points between days 8 and 45 of differentiation for scRNA-seq (Note S1). Interestingly, at day 25, a population of PSC-CMs emerged with light-scattering properties that seemed to indicate larger cells (Figure S3A; supplemental methods). Based on this, we sorted both “large” and “normal” size populations at days 25, 30, and 45. We sequenced 660 PSC-CMs to a median depth of ~11,400 UMIs per cell and ~3,000 genes per cell. SingleCellNet further validated the cells as having CM identity (Figure S1A).
As with endogenous CMs, Monocle 3 recovered a unidirectional, non-branching trajectory for PSC-CMs (Figure 2C; Note S3). Pseudotime progressively but heterogeneously increased over biological time, with a general transition occurring between days 10 and 25 of differentiation (Figure 2D). These general dynamics were further validated by RNA velocity analysis (Figures S3B and S3C). We observed some, albeit minimal, line-to-line differences in pseudotime progression (Figure S3D). Interestingly, there appeared to be no major pseudotime differences between the identified “large” and “normal” cells (Figure S3E). This may potentially indicate phenotypic differences in scatter properties that are mediated by non-transcriptional mechanisms. Nevertheless, given these results, we treated these cells as identical for our downstream analyses. We identified 449 differentially expressed genes over the PSC-CM trajectory. These genes were enriched for terms related to cardiac structure, contractile function, electrophysiology, and metabolism (Figure 2E). Taken together, these results validate the successful reconstruction of a PSC-CM maturation trajectory.

**PSC-CMs are transcriptional nearest neighbors to embryonic/neonatal CMs**

Having reconstructed trajectories for both in vivo and in vitro CM maturation, we next sought to align the trajectories to determine how PSC-CMs compared with their in vivo counterparts. We combined the two sets of cells and performed trajectory reconstruction (Note S4). To our surprise, the in vivo CM and PSC-CM trajectories remained completely separate and failed to align (Figure 2F). This initial result suggested significant global gene expression differences between the two groups of CMs, despite common genetic background.

To overcome this issue, we made use of properties of mnnCorrect, which is the default batch correction algorithm implemented in Monocle 3. mnnCorrect projects cells onto a reference batch by identifying mutual nearest neighbors between the batches and correcting differences between mutual nearest-neighbor pairs. We used mnnCorrect to align the in vivo and in vitro trajectories in two separate but similar approaches. In the first approach, we set the group (e.g., in vivo/in vitro) to be a batch effect to be corrected (Figures 2G and 2H); thus, mnnCorrect explicitly looked for mutual nearest neighbors between the groups of cells. In the second approach, we set the reference batch to be a batch containing only in vivo CMs (Figures 2I and 2J). This implicitly treats the group as a batch effect by again forcing PSC-CMs to be projected onto in vivo CMs. Both approaches make the assumption that group differences are either driven by technical artifacts or by biological phenomena that are not of interest. This assumption is likely to be incorrect—indeed, global expression differences may be directly biologically relevant to the poor maturation status of PSC-CMs (which we consider below). Nevertheless, this assumption was appropriate for the initial goal of identifying an approximate alignment between the endogenous CM and PSC-CM trajectories.

Both approaches yielded successful integration of the in vitro and in vivo CM trajectories (Figures 2G–2J). Notably, in both approaches, the vast majority of PSC-CMs showed pseudotime scores similar to endogenous CMs from E18 to P4 (e.g., the embryonic/neonatal period). While some PSC-CMs appeared to enter into the perinatal phase of maturation,
nearly no PSC-CMs successfully achieved an adult phenotype. This result is in line with the hypothesis that the previously identified perinatal phase of CM maturation is somehow disrupted in PSC-CMs, leading to their immature phenotype.

As a complementary approach to trajectory reconstruction, we also used our previously developed transcriptomic entropy metric\(^5\) to stage the PSC-CMs. This metric is based on the observation that immature cells display a broader, more promiscuous gene expression profile, which subsequently narrows as the cells mature. Thus, immature CMs will display a high transcriptomic entropy, while mature cells will show a lower transcriptomic entropy. Our previous work optimized this metric to enable direct comparisons across multiple datasets while being robust to technical batch effects. Here, transcriptomic entropy of the PSC-CMs was generally high and corresponded to that of embryonic/neonatal CMs, with very few PSC-CMs demonstrating lower transcriptomic entropy than P8 CMs (Figure 2K). These results provide a trajectory reconstruction-independent validation of our above results. Additionally, these results corresponded well with our previous analysis of transcriptomic entropy in human PSC-CMs. Taken together, our data support the embryonic/neonatal maturation status of PSC-CMs.

**PSC-CMs show global expression differences from endogenous CMs**

As a first step in understanding the immature phenotype of PSC-CMs, we sought to determine the global gene expression differences that led to complete separation of the in vivo and in vitro trajectories. To this end, we tested for differential gene expression between all in vivo CMs and PSC-CMs (Figure 3A). This identified 2,906 differentially expressed genes. 1,460/2,628 (56%) of our previously identified maturation genes showed global gene expression differences, suggesting significant differences between in vivo and in vitro CMs. However, one explanation for this could be that this comparison includes mature endogenous CMs, and thus we are simply capturing the immaturity of PSC-CMs. We therefore tested for differential gene expression between early-stage CMs (defined as in vivo pseudotime\(\in [0, 15]\)) and PSC-CMs (Figure 3A). This yielded 1,743 differentially expressed genes, with 1,017/2,628 (39%) maturation genes falling into this category. Thus, PSC-CMs showed global expression differences even when compared with endogenous CMs identified as their transcriptional nearest neighbors.

Gene Ontology analysis of globally differentially regulated genes identified terms corresponding to CM maturation (Figure S4A). Genes associated with sarcomeric structure, oxidative phosphorylation, fatty acid metabolism, and calcium handling were generally expressed at higher levels in vivo, while genes associated with proliferation/cell cycle, stemness, and transcription/translation were higher in PSC-CMs. We further investigated expression differences in major sarcomeric, mitochondrial, and ribosomal genes (Figure 3B). In general, sarcomeric proteins and mitochondrial-related transcripts (e.g., mitochondrially encoded proteins, electron transport proteins, and ATPases) were more highly expressed in vivo, while ribosomal protein-coding transcripts were more highly expressed in PSC-CMs. Notably, while PSC-CMs expressed both ventricular myosin light chain genes (\(\text{Myl2, Myl3}\)) and atrial genes (\(\text{Myl4, Myl7}\)), they expressed the former pair at a much lower level than in vivo CMs and the latter pair at a much higher level. This is complicated by the fact that
endogenous ventricular CMs express Myl4 and Myl7 during embryonic stage. Thus, these results may indicate a more “atrial-like” phenotype for PSC-CMs, but they may also reflect a general inadequacy in using these markers to classify CMs. Additionally, PSC-CMs showed comparatively higher expression of the immature troponin I isoform Tnni1 and lower expression of the mature isoform Tnni3.

On the surface, these results seem to reiterate the immature phenotype of PSC-CMs. However, we observed that PSC-CMs do not just demonstrate immature-like gene expression levels. Rather, their absolute gene levels often fall entirely outside the spectrum of endogenous development. To quantify this, we scaled all of the maturation genes such that 0 corresponded to the lowest level over in vivo maturation, while 1 corresponded to the highest level. We then normalized the average PSC-CM gene expression of the maturation genes by this same scaling. Notably, 1,186/2,628 genes (45%) fell outside the range of [0, 1] (Figure 3C). Strangely, many genes showed lower overall expression levels in PSC-CMs, even if those genes were downregulated during endogenous maturation. Gene expression levels for in vivo and in vitro CMs may thus fall into two entirely different biological scales. These results highlight the challenge of comparing between the in vivo and in vitro contexts — direct comparison of absolute gene levels may be difficult or even impossible. We discuss these issues in more detail in Note S5.

**PSC-CMs poorly recapitulate perinatal CM maturation-related gene changes**

As expression levels may be on differently calibrated biological scales for in vivo and in vitro CMs, direct comparison of expression levels may not yield insights into the dysregulation of maturation in vitro. Thus, we proceeded with a different approach by drawing insights from a previous study from our group. There, we sought to compare first and second heart field progenitor cells isolated from in vivo embryos and in vitro precardiac organoids by RNA-seq. While absolute expression levels corresponded poorly between the in vivo and in vitro samples, the directionality of differences in marker genes between first and second heart field cells was well preserved. Applying that lesson here, rather than viewing the absolute gene expression levels of the adult CM as the gold standard, we focused instead on the directionality of gene changes from immature to mature CMs. We could then see how well these relative changes were recapitulated in PSC-CMs. This approach enabled comparison between the in vivo and in vitro CM maturation trajectories without the need to directly compare expression levels.

We first computed the FCs for all of the maturation genes across the in vivo trajectory, both in the pseudotime interval [0, 20] (encompassing the embryonic/neonatal period) and in the interval [0, 40] (encompassing both the embryonic/neonatal and perinatal periods). We then compared these against the FCs of the maturation genes across the in vitro trajectory. Notably, for genes that were differentially expressed both in vivo and in vitro, FCs were generally comparable (Figure 3D). However, only 10% of maturation genes were differentially expressed in the same direction in PSC-CMs (Figure 3E). These data support the failure of PSC-CMs to successfully recapitulate the majority of maturation-related gene changes.
We next used our approach from Figure 1D to determine which maturation genes are best recapitulated in vitro. We quantified what percentage of differentially expressed genes within each pseudotime subset are also differentially expressed in the same direction in vitro. While recapitulation in generally poor, genes that become differentially expressed during the embryonic/neonatal period were more likely to be correctly differentially expressed in vitro (Figure 3F). For example, 37% of genes differentially expressed from [0, 10] in vivo are correctly differentially expressed in vitro, and 28% of genes differentially expressed from [0, 20] in vivo are captured in vitro. However, this number drops to 11% for genes differentially expressed from [0, 40]. The sharp drop-off after in vivo pseudotime 20 (corresponding to the start of the perinatal phase) demonstrates that PSC-CM maturation is particularly poor at recapitulating perinatal maturation-related changes.

As a further step, we quantified what percentage of genes in each of the identified in vivo clusters are correctly recapitulated in vitro. By far, genes from EA1 are best recapitulated in PSC-CMs, while less than 10% of genes in EA2, EA3, and LA1 are correctly differentially expressed in vitro (Figure 3G). Most strikingly, only 1.1% of LA2 genes are captured in vitro. In summary, PSC-CMs undergo a subset of embryonic/neonatal-associated maturation changes but fail to undergo the perinatal program. This may in turn point to their phenotypic maturation arrest.

Human PSC-CMs show poor recapitulation of perinatal CM maturation-related gene changes

Given the importance of PSC-CM technology for applications in human health, we next investigated maturation-related changes in human PSC-CMs. In our previous study, we found that, similar to mouse PSC-CMs, human PSC-CMs appear to have a maturation state similar to fetal CMs and appear arrested at the onset of perinatal maturation. Here, we focused on four published datasets of CMs generated from induced PSCs—Friedman, Nguyen, and Lukowski et al. (FNL); Churko et al. (C); Gerbin, Grancharova, Donovan-Maye, and Hendershott et al. (GGDH); and Ruan and Liao et al. (RL). We initially aimed to perform trajectory reconstruction with Monocle 3 as done with the mouse PSC-CMs. However, none of the datasets formed a smooth, continuous trajectory but rather showed discrete separation of time points (Figure S4B). In each of these datasets, samples from individual time points were typically prepared as separate batches, and thus batch and time point were confounded. Thus, it was difficult to resolve potential batch effects to create an appropriate trajectory. As a workaround, we applied our transcriptomic entropy approach, which we previously showed can also function as a surrogate pseudotime. Transcriptomic entropy largely recapitulates similar differentially expressed genes as commonly used trajectory inference methods while being generally resistant to batch effects. Through this method, we identified differentially expressed genes for each of the four datasets.

As with the mouse PSC-CMs, we compared the directionality of changes in maturation genes for the human PSC-CMs against our in vivo data. This approach fundamentally assumes that maturation-related changes are comparable across human and mouse. Whether this universally holds requires further assessment; however, in the absence of comprehensive perinatal human CM scRNA-seq data, our assumption served as a useful first approximation.
We found that while several of the human datasets performed better than the mouse PSC-CMs in recapitulation of maturation genes, all showed a sharp drop-off in correct recapitulation at the onset of the perinatal period (Figure 3H). Likewise, while the datasets performed better in terms of recapitulating genes in the EA2 and EA3 clusters, the LA1 and LA2 clusters were still relatively poorly captured (Figure 3I). The one exception to this observation was the GGDH dataset, which appeared to perform poorly in general, though this may also be for technical reasons owing to the lower depth/sensitivity of that study. As a whole, however, the data indicate that human PSC-CMs similarly fail to undergo perinatal programs associated with CM maturation.

A network of nine TFs underlies dysregulation of PSC-CM maturation

Given that the poor PSC-CM maturation phenotype can be seen across multiple lines and protocols from several species, we hypothesized that there is a conserved mechanism underlying maturation failure in PSC-CMs. Commonly dysregulated genes across multiple studies may point to a source for disruption of perinatal maturation programs. We thus identified dysregulated genes for the mouse PSC-CMs generated in this study as well as the four literature-obtained human PSC-CM datasets (Figure 4A). We used the following criteria to classify a gene as dysregulated: either upregulated in vivo but not in vitro or down-regulated in vivo but not in vitro and expressed in vitro. The latter criterion allowed us to eliminate genes downregulated in vivo but already not expressed in vitro, as these genes are less likely to be relevant to maturation failure. We generated a consensus human list by including genes dysregulated in at least three of the four literature datasets. This list was intersected with the dysregulated genes from the mouse dataset to generate a final list of 550 genes dysregulated across the studies and across species.

The consensus dysregulated gene list was notably enriched for genes differentially expressed during the perinatal period in vivo (Figure 4B). Similarly, the time to 10% FC, time to 50% FC, and time to 95% were all higher for the dysregulated genes compared with correctly recapitulated genes (Figure S4C). Lastly, the consensus dysregulated gene list was enriched for genes from the LA clusters, with 33% and 42% of genes from LA1 and LA2 dysregulated, respectively. These results are in line with our hypothesis that the perinatal period of maturation is particularly disrupted in PSC-CMs. We additionally investigated the chromatin accessibility of the dysregulated genes compared with correctly recapitulated genes by analysis of three previously published assay for transposase-accessible chromatin (ATAC)-seq datasets of human PSC-CMs at days 15, 25, and 30. The percentage of dysregulated genes with peaks in the promoter-transcription start site (TSS) region was marginally lower than for correctly recapitulated genes (67% vs. 72%) (Figure S4D); however, this difference is unlikely to be biologically relevant. This suggests that dysregulated genes show a similar level of chromatin accessibility to correctly recapitulated genes.

Given the similarity in chromatin accessibility, we next aimed to identify upstream factors that could be responsible for gene dysregulation in PSC-CMs. As before, we performed over-representation analysis to identify TFs whose targets are particularly highly represented in the consensus dysregulated gene list. Using affinity propagation to eliminate redundancy...
(STAR Methods), we narrowed down the identified TFs to a list of nine candidate TFs (Figure 4C). We refer to these TFs from here as dysregulated maturation TFs. Many of the dysregulated maturation TFs were identified as being important regulators of in vivo CM maturation in our above analysis (Figure 1K). Additionally, almost all have been previously directly implicated in CM differentiation, maturation, or disease response.11,29–32,45–51 The STRING protein database identified significant connectivity between these TFs (Figure S4E), with a protein-protein interaction enrichment p value of $1.75 \times 10^{-8}$. Thus, the identified TFs likely work as a regulatory network to mediate CM maturation; disruption of this network may underlie maturation failure in vitro.

The disruption of our identified TF network could occur at multiple stages, including at the gene expression, protein level, or protein activity level. As a first step, we investigated the transcriptional levels of each of the dysregulated maturation TFs in our paired in vivo and in vitro CMs (Figure 4E). In vivo, all were expressed early in maturation. Subsequently, some decreased in level over the maturation process (Yy1, Jun, Nfe2l2, Sox9, Nrf1, Mef2a), while others remained expressed at a relatively constant level (Srf, Ppara, Essra). This suggests that these TFs may play a role in triggering the maturation process but are dispensable afterward. By contrast, in PSC-CMs, nearly all showed much lower levels compared with in vivo CMs, particularly at the start of in vitro maturation. This supports the possibility of network failure at the gene expression stage. Notably, PSC-CMs did express similar transcriptional levels to in vivo CMs for regulators of the EA1 cluster (particularly Elk1 and the HSF and GABP genes), supporting failure of the perinatal program specifically (Figure S4F). However, as discussed earlier, it is also possible that in vivo and in vitro CMs have different scales of expression, complicating direct comparison of expression levels. As an alternate comparison, we used Ingenuity Pathway Analysis (IPA) to infer TF activity. IPA infers activity based on FCs of downstream genes compared against known literature interactions.52 Using IPA, we found that all dysregulated maturation TFs, with the exception of PPARA, showed either weaker or reversed activity over PSC-CM maturation compared with in vivo maturation (Figure 4F). Our results support the hypothesis that these dysregulated maturation TFs play a role in the failure of PSC-CMs to undergo perinatal maturation programs.

**Ex vivo perturbations only partially activated dysregulated maturation TF network**

A number of cellular and tissue engineering methods have been proposed to improve PSC-CM maturation. However, in the absence of knowledge about in vivo maturation or the nature of PSC-CM maturation dysregulation, it has been difficult to assess how biomimetic these methods are. In particular, while these methods may impact some functional characteristics associated with maturing CMs, the mechanism of these changes may be different than endogenous maturation. Our findings here provide a useful for framework for investigating the effects of ex vivo perturbations. If a particular perturbation indeed improves PSC-CM maturation in a biomimetic manner, then the direction of gene changes between perturbed vs. control tissue should match the direction of changes during in vivo maturation.
To this end, we identified nine publicly available RNA-seq datasets spanning multiple different perturbations of PSC-CMs\textsuperscript{53–61} (Figures 5A and S5). For each, we identified differentially expressed genes between the provided experimental and control groups and intersected these genes against our identified maturation genes (Figures 5B and S5A). Notably, for most of the perturbations, a majority of the identified differentially expressed genes matched the direction of \textit{in vivo} CM maturation. However, a large number (average of \textasciitilde38\% of genes across the studies) were differentially expressed in the opposite direction. While these may represent genes further dysregulated by perturbations, they may also reflect differences in species as well as differences in technical conditions across all of the different studies (e.g., time points, isolation protocols, degree of cellular purity). Thus, we focused our analysis on genes that were differentially regulated in the same direction between perturbations and \textit{in vivo} maturation. Of these genes, we found that \textasciitilde30\%–50\% were already correctly differentially expressed over control PSC-CM maturation in our previously identified datasets (Figure 5C). However, \textasciitilde10\%–25\% of differentially expressed genes came from our consensus dysregulated gene list. Thus, while perturbation methods predominantly impact genes that are already being correctly differentially expressed \textit{in vitro}, they may also “correct” previously dysregulated genes.

For most perturbations, the identified differentially regulated genes were typically enriched in the embryonic/neonatal time points (Figures 5D and S5B), and enriched in genes from EA1 (Figures 5E and S5C). This suggests that most perturbations predominantly continue to activate early-stage gene changes without affecting the dysregulated perinatal phase. However, some perturbations proved to be exceptions to this general rule. In particular, long-term culture (LTC), engineered heart or microtissues (EHTs/EMTs), and treatment with PPAR\(\gamma\) activator appeared to show prominent enrichment of LA1 and LA2 compared with EA1 within the same study, potentially indicating that these methods may be better at improving maturation in PSC-CMs.

To further investigate how perturbations affect PSC-CM dysregulation, we looked at changes to our identified dysregulated maturation TFs. We looked at two parameters—whether the TF itself was upregulated by the perturbation (Figures 5F and S5D) as well as whether downstream targets of the TF were enriched in the differentially expressed genes of that perturbation (Figures 5G and S5D). The two methods show decent, but not complete, overlap in results. This may indicate that some methods work through posttranscriptional methods of TF activation, though they may also reflect differences in underlying computational assumptions between the methods. Notably, electrical stimulation (ES) showed the most prominent transcription-level changes to dysregulated maturation TFs, while LTC, EHTs/EMTs, C16 media, and ERR\(\gamma\) activation showed the largest effect on downstream targets. However, while each method was able to activate some of the dysregulated TFs, no method could completely activate them all. This suggests an inherent limitation in the ability of current \textit{ex vivo} approaches to fully overcome the PSC-CM maturation deficit.
**DISCUSSION**

In this study, we used scRNA-seq to identify developmental processes associated with CM maturation. In particular, for the first time, we reconstructed a high-quality trajectory of *in vivo* CM maturation with significant sampling of perinatal time points. While gene trends associated with CM maturation initiate early (~E18.5), CMs undergo a perinatal phase (P8–P15) during which the rate of transcriptional changes is at its highest and most genes progress to their mature levels. Through comparison of the trajectories, we found that this perinatal phase is largely not recapitulated in PSC-CMs. We identified a network of nine TFs upstream of dysregulated genes in PSC-CMs; these TFs consistently showed lower expression and disrupted activity *in vitro*. Our study thus provides a transcriptional underpinning for maturation programs in endogenous CMs and how this process may be disrupted in PSC-CMs, leading to a poor maturation state.

Additionally, by surveying published datasets of *ex vivo* perturbations, we found that no method can fully activate all of the identified TFs. To date, the mechanisms by which *ex vivo* perturbations affect CM biology is often unclear, and maturation is often assessed based on *ad hoc* measurements. Thus, it is possible that these methods work through non-developmentally mimetic mechanisms. We believe that for optimal engineering of PSC-CMs, *in vivo* development must be viewed as the gold standard. Our results emphasize the importance of comparing perturbations directly to endogenous CMs to establish claims of cellular maturation.

CM maturation continues to remain a somewhat poorly understood phenomenon. For example, it is not fully clear what initially triggers maturation. Our trajectory reconstruction indicates that maturation processes begin *in utero*, in line with previous findings. Thus, it is unlikely that birth, and its accompanying hemodynamic and metabolic changes, is the initial driver of CM maturation, though these changes may play a subsequent role. The role of various neuroendocrine cues must be further studied here, as glucocorticoids, thyroid hormone, IGF1, and NRG1 all show spikes in the late embryonic period. Likewise, while we identified TFs that regulate various clusters of genes through embryonic and perinatal maturation, it is not clear what stimulates activation of these TFs.

One interesting question is if there is some type of stepwise activation of the identified maturation TFs, perhaps corresponding with sequential changes in various individual aspects of maturation (e.g., hypertrophy, sarcomeric organization, electro-physiological changes, metabolic adaptations). For example, a recent study found evidence that sarcomere maturation may be upstream of and regulate hypertrophy, mitochondrial remodeling, and t-tubule formation. The analysis of *ex vivo* perturbations in this study and others, however, suggests more interconnected processes. For example, perturbations like stretch, ES, anisotropic patterning, and others activate maturation-related TFs in a partially, but not completely, overlapping fashion. Thus, maturation TFs may both regulate and be regulated by biophysical changes occurring in the pre- and postnatal heart. Future studies can better clarify whether there are stage-specific regulatory aspects to maturation, which may also affect the optimal protocols for maturing PSC-CMs.

*Cell Rep. Author manuscript; available in PMC 2023 October 23.*
Similarly interesting is the overlap in upstream TFs across gene clusters with different temporal dynamics. Future studies should more thoroughly identify how common TFs can mediate different expression dynamics, for example through differences in chromatin accessibility, binding affinity, or co-localized TFs.

We identified nine TFs that are likely to underlie maturation failure in PSC-CMs. Interestingly, the dysregulated targets of these TFs show similar chromatin accessibility compared with correctly differentially expressed genes. However, the expression levels of each of the maturation TFs is much lower in vitro than in vivo. Thus, correcting the expression levels of these TFs provides a putative target for future genetic engineering efforts to improve PSC-CM maturation. Intriguingly, these TFs display significant known interactions with one another, potentially suggestive of a regulatory network for CM maturation. One consequence of this observation is that it is unlikely that upregulation of any one TF alone will induce maturation. Indeed, a recent study found that upregulation of SRF in neonatal CMs led to disruption of maturation. Thus, balanced and controlled gene dosages are likely necessary for these TFs to effect maturation. We anticipate that a reprogramming-type approach, featuring concomitant upregulation of multiple dysregulated maturation TFs, will enable future generation of mature PSC-CMs.

**Limitations of the study**

In this study, we generated references of in vivo and in vitro CM maturation that we aimed to be readily usable by others studying cardiac biology. Given our focus on understanding gene regulatory pathways in maturation, we focused on generating a high-quality reference by combining LP-FACS with a high-sensitivity plate-based sequencing protocol (mcSCRB-seq) and relatively high sequencing depth. As a trade-off, however, we sequenced relatively fewer cells (~50–100/time point). While we believe our reference should be adequately powered for trajectory reconstruction (Note S1), further studies will certainly benefit from increased cell numbers for more robust reconstruction and improved biological discovery. LP-FACS can be readily integrated with higher-throughput sequencing protocols (such as sci-RNA-seq) to achieve this need in the future.

Our findings here are based on several assumptions that must be considered critically. In comparing the in vivo and in vitro trajectories, a fundamental assumption is that all gene changes that occur over endogenous maturation must be associated with or required for completion of maturation. However, many identified differentially expressed genes may be inessential. Indeed, we were surprised to note that many downregulated genes in vivo already present with a lower expression in vitro. Though we filtered many of these genes from our identified consensus dysregulated list, other genes may similarly be dispensable for maturation. While our study is not equipped to identify such genes, the emergence of improved screening methodologies may help better refine the core gene dynamics of CM maturation in the future.

A second assumption was that gene trends across maturation should be generally comparable across multiple datasets. This assumption is reasonable for comparing mouse endogenous CMs and PSC-CMs generated within the same study, as done here. However, whether this assumption holds for our various meta-analyses must be considered more
carefully. For one, it is not fully clear whether CM maturation trends must be preserved across species. While recent results from Cardoso-Moreira indicate that developmental trends between species are often divergent, the results from Uosaki et al. suggest that CM maturation in particular is relatively well preserved. Both sets of studies were done using bulk cardiac data, and scRNA-seq data are currently unavailable for human perinatal time points. Thus, while our approach functions as a useful first approximation of studying PSC-CM dysregulation, there is also a need for better-resolution human data to further validate our findings. This is particularly important given the need for human PSC-CMs for clinical application. Likewise, there are generally few well-controlled scRNA-seq studies of ex vivo perturbations to improve CM maturation. We anticipate that future datasets can provide better insight into the molecular pathways by which specific perturbations impact CM maturation.

Thirdly, our study primarily focused on transcriptional mechanisms of CM maturation and PSC-CM dysregulation. However, there is plenty of evidence to implicate posttranscriptional processes in CM maturation. For one, we observed that CMs achieve a largely matured transcriptome by P15–P18. However, protein-level changes continue to occur after this time, and CMs do not achieve their maximal volume until approximately 3 months of age in mice. Moreover, protein-protein interactions at the cell membrane and with the extracellular matrix mediate important changes in postnatal CM biology. We expect that the emergence of improved proteomics methods, including single-cell proteomics, will enable better resolution of posttranscriptional maturation processes. Nevertheless, our results make it clear that CM maturation is associated with large-scale transcriptional changes and that these changes are not fully recapitulated in vitro. Thus, it is likely that transcriptional mechanisms are necessary, though perhaps not sufficient, for successful CM maturation.

Despite these assumptions, we believe this study takes the first steps toward understanding the nature of PSC-CM maturation failure by direct comparison to endogenous developmental processes. These findings can serve as a launch point for future efforts to improve the clinical applicability of PSC-CMs.

**STAR METHODS**

**RESOURCE AVAILABILITY**

**Lead contact**—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Chulan Kwon (ckwon13@jhmi.edu).

**Materials availability**—Four mESC lines (aMHC-cre x mTmG) were generated as part of this project. We are happy to share lines on request pending appropriate material transfer protocols.

**Data and code availability**—All of the sequencing data for this study can be found on GEO with accession number GSE164591. Additionally, the code to reproduce all of the figures in the manuscript is available on Github at https://github.com/skannan4/cm-dysregulation (Zenodo https://doi.org/10.5281/zenodo.7703280). Lastly, we have made an
R workspace available on Synapse ([https://www.synapse.org/#!Synapse:syn23667436/files/](https://www.synapse.org/#!Synapse:syn23667436/files/)) that contains many of the data tables pre-loaded for our analysis. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

**EXPERIMENTAL MODELS AND SUBJECT DETAILS**

**Mouse lines**—To generate mice for our reference dataset, we crossed B6.FVB-Tg(Myh6-cre)2182 Md/J mice (aMHC-cre, Jackson Laboratory, Stock No. 011038) with B6.129(Cg)-Gt(Rosa)26Sor1M4(TdTomato,-EGFP)Luo/J (mTmG, Jackson Laboratory, Stock No. 007676). Both mice have C57BL/6J congenic background. A combination of male and female mice were used to capture timepoints spanning from e14.5-p84 per the manuscript. All animals were maintained compliant to protocols by the Johns Hopkins Animal Care and Use Committee.

**Embryonic stem cell lines**—To generate mESCs for *in vitro* studies, mice of the above strain were crossed and observed for the presence of a vaginal plug (considered embryonic day 0.5). On embryonic day 3.5, blastocysts were flushed, isolated, and maintained in pre-gelatinized 96 well plates with 2i media. Four genotype-verified mESC lines were expanded and tested for differentiation capability to the cardiomyocyte lineage.

**METHOD DETAILS**

To generate mice for our reference dataset, we crossed B6.FVB-Tg(Myh6-cre)2182 Md/J mice (aMHC-cre, Jackson Laboratory, Stock No. 011038) with B6.129(Cg)-Gt(Rosa)26Sor1M4(TdTomato,-EGFP)Luo/J (mTmG, Jackson Laboratory, Stock No. 007676). Both mice have C57BL/6J congenic background. To generate mESCs for *in vitro* studies, mice of the above strain were crossed and observed for the presence of a vaginal plug (considered embryonic day 0.5). On embryonic day 3.5, blastocysts were flushed, isolated, and maintained in pre-gelatinized 96 well plates with 2i media. Four genotype-verified mESC lines were expanded and tested for differentiation capability to the cardiomyocyte lineage. All animals were maintained compliant to protocols by the Johns Hopkins Animal Care and Use Committee.

**CM isolation**—For isolation of CMs from e14-p4 timepoints, we used the neonatal cardiomyocyte isolation kit from Miltenyi Biotec in conjunction with the gentleMACS Dissociator. For later timepoints, we performed Langendorff isolation of CMs. We prepared the following buffers:

- **Perfusion buffer:** 120 mM NaCl, 5.4 mM KCl, 1.2 mM NaH₂PO₄, 20 mM NaHCO₃, 5.5 mM glucose, 5 mM BDM, 5 mM Taurine, and 1 mM MgCl₂, adjusted to pH 7.4
- **Digestion buffer:** 40 mL Perfusion buffer plus 35.8 mg Collagenase Type II (Worthington CLS-2), 3 mg Protease (Sigma P5147)
- **Tyrode’s buffer:** 140 mM NaCl, 5 mM KCl, 10 mM HEPES, 5.5 mM glucose, and 1 mM MgCl₂, adjusted to pH 7.4
We used a horizontal (i.e. non-hanging) Langendorff apparatus with a chamber filled with perfusion buffer. To perform isolation, we first performed isofluorane anaesthesia on non-heparinized mice. Mice were observed until clearly anaesthetized and unresponsive to toe pinch, and subsequently euthanized by cervical dislocation. The heart was then rapidly excised from the chest and cannulated to the Langendorff apparatus. Flow time and rate of flow were dependent on the age of the mouse and were typically judged based on completeness of digestion to touch. Subsequently, the left ventricular free wall was excised and minced. We filtered isolated cells through a 100 μM screen to eliminate large tissue chunks, spun down at 800 RPM for 1 minute (Eppendorf centrifuge 5702), and resuspended cells in 10 mL Tyrode’s buffer.

PSC-CM differentiation—Our PSC-CM differentiation protocol was adapted from multiple previously published protocols. The day prior to initiation of differentiation protocol (D-1), cells were changed to expansion media with 1000 U Lif/mL with no CHIR99021 or PD0325901. On Day 0 of differentiation, cells were dissociated with TrypLE (Thermo Fisher) and suspended in a serum-free differentiation media (SFD). SFD was composed of ¼ volume IMDM to ¼ volume Ham’s F12 media, with 0.5% v/v N2 supplement (Gibco), 1% v/v B27 without retinoic acid (Gibco), 0.5% v/v BSA (Sigma) in PBS, 0.75% v/v glutamine (Gibco), 0.75% v/v penicillin-streptomycin (Gibco), 50 μg/mL ascorbic acid (Sigma), and 0.039 ul/mL 1-thioglycerol. For the first four days of differentiation, differentiating cells were maintained as embryoid bodies. On Day 2, media was replaced with fresh SFD plus 3uM CHIR99021 (Selleckchem) and 2.5 ng/mL BMP4 (R&D Systems). On Day 4, cells were dissociated with TrypLE and replated as a confluent monolayer on gelatin-coated flasks. At this time, cells were cultured with SFD with 0.1% v/v XAV939. On Day 6, the media was changed to SFD. From Day 10 to Day 13, lactate selection was performed by culturing the cells in DMEM without glucose plus lactate. Subsequently, the cells were cultured in SFD, with media changes every two days.

FACS for single CM and PSC-CM isolation—For isolating endogenous CMs, we used LP-FACS. We have detailed our LP-FACS approach previously.19 We reproduce our methods here. We utilized a COPAS SELECT instrument (Union Biometrica). The COPAS SELECT was updated and rebranded as the FP-500, but the protocol here study does not use the new features and thus the two are functionally indistinguishable. We optimized sorting for cardiomyocytes by using a sort delay of 8 and sort width of 6. Additionally, we used the following fluorescence settings: ext gain 50, green gain 200, yellow gain 200, red gain 255, extension integral gain 50, green integral gain 200, yellow integral gain 200, red integral gain 255, green PMT 800, yellow PMT 800, red PMT 1100. Coincidence check was selected to ensure proper single event sorting. We typically flowed cells between 20 – 60 events/second. We maintained cells in Tyrode’s buffer during the sort and sorted them into prepared collection plates for mcSCRB-seq library prep. To run the machine, we used ClearSort Sheath Fluid (Sony, Lot 1218L345).

PSC-CMs, unlike endogenous CMs, do not retain their shape when dissociated and instead round up. Thus, they fall below the recommended range of sorting through LP-FACS, but can be readily sorted through conventional FACS. Here, PSC-CMs were dissociated from
culture flasks with TrypLE (5-10 minutes, depending on the timepoint). Cells were strained to remove clumps and clusters and subsequently sorted into prepared collection plates on either a MoFlo Legacy or MoFlo XDP. We isolated healthy singlets by first gating on forward and side scatter, followed by forward scatter and pulse width. We used a propidium iodide stain for further isolation of healthy cells, sorting out GFP+/PI− cells. Interestingly, from D25 onwards, we observed that GFP+/PI− cells appeared to split into two populations based on PI autofluorescence (despite being PI−). When analyzed, the high autofluorescence cells also appeared to have higher side scatter and pulse width (Figure S3A). Upon sorting, these cells also appeared to be visually larger under the microscope. Given this, we sorted cells from both the “normal” and “larger” populations at D25, D30, and D45.

**scRNA-seq library preparation and sequencing**—We performed sequencing using the mcSCRB-seq protocol. The protocol has been described at protocols.io at dx.doi.org/10.17504/protocols.io.p9kdr4w. Information about the library design is provided in Methods S1; more detailed metadata is also available on Synapse. We sequenced the final, pooled library on a NovaSeqS4 as 150-base pair paired-end reads. We subsequently demultiplexed into two files such that the read 1 file contains the 8 base pair i7 tag, 6 base pair cell barcode, and 8 base pair UMI; read 2 contains the 150 base pair cDNA read. We have provided these final demultiplexed reads to GEO at accession GSE164591. However, if the original paired-end reads from the sequencer are desired, we are happy to provide on request. Mapping of the data was done with kallisto|bustools (0.46.2), using an index generated from the CellRanger mouse reference concatenated with the ERCC spike-in sequences. For RNA velocity, mapping was done with kallisto|bustools using special indices with intronic and exonic sequences respectively from GRCm38.98.

**QUANTIFICATION AND STATISTICAL ANALYSES**

Most analyses performed in the paper were done in R 4.0.x-4.1.x (with the exception of RNA velocity, done in Python); code to reproduce the figures can be found at our Github (https://github.com/skannan4/cm-dysregulation). We encourage readers to look directly to the code for specific technical details about our method, and we are of course happy to answer additional questions on request. However, here, we briefly annotate methods used throughout the manuscript. Further details regarding specifics of trajectory reconstruction can be found in Methods S1.

**General quality control**—Quality control continues to be a major issue in scRNA-seq analysis. Poor quality cells can confound analyses and need to be removed. In this manuscript, we used the general approach to quality control established in our previous work. We used three parameters: percent of reads going to the top 5 genes, depth, and CM identity. For the first two parameters, we normalized the computed metric against the median value for that timepoint, since both metrics will inherently vary as CMs mature. Our thresholds were top5_norm < 1.8 and depth_norm > −0.7. These are somewhat more permissive than the thresholds we set in our previous work; however, in that study, we were working with many datasets, including droplet datasets were poor quality cells are more abundant. We found empirically that these relaxed thresholds were sufficient for our higher quality plate-based data. For CM identity, we used the singleCellNet package (0.1.0),21
compared against the Tabula Muris.\textsuperscript{77} We selected all cells whose highest scoring identity was "cardiac muscle cell." See Figure S1A.

**Trajectory reconstruction**—For trajectory reconstruction (Figures 1B, 1C, 2C, 2F–2K), we used Monocle 3 (0.2.3.3).\textsuperscript{23} Monocle 3 uses fastMNN based on the batchelor package (1.2.4), which in turn is based on mnnCorrect.\textsuperscript{36} We discuss some specifics of trajectory reconstruction (e.g. number of principal components used, batch effect removal) in the Supplementary Notes and Figure S6. More specific details such as trajectory start points and graph settings can be found in the code.

**Differential gene expression and gene Ontology analysis**—Differential gene expression analysis for single cell datasets (Figures 1D, 2E, 3A–3I, 5A–5E, S1B, S2A–S2D, S4A and S4B) was done in Monocle 3. We typically set a cutoff such that testing was only done in genes expressed in at least 25\% of cells; we used this as our cutoff for whether a gene was “expressed” in a given cell group. A Benjamini-Hochberg-adjusted p-value threshold of \( q < 0.05 \) was used to determine significance.

For bulk datasets, as in Figure 5, we instead used DESeq2 (1.26.0),\textsuperscript{76} typically setting the design to compare between perturbation and control. We used the Benjamini-Hochberg-adjusted p-value threshold of \( q < 0.05 \) to determine significance. However, because bulk samples can detect genes with higher sensitivity, we additionally used a fold change threshold of \( |\log_2(FC)| \geq 0.5 \).

For Gene Ontology analysis, we predominantly used the resource at the Gene Ontology website (\texttt{http://geneontology.org/}), which in turn links to the PANTHER classification system.\textsuperscript{78} To more readily visualize Gene Ontology terms while removing redundancy, we selected the top 150 terms by enrichment and input into REVIGO.\textsuperscript{79} The exception to this workflow was in Figure S2, where we instead used WebGestalt 2019 with the following settings: biological processes terms, size limit 850, top 25 terms, weighted cover set expecting 10. Our rationale for using this second method was that the terms provided were more readily condensed for easier broad visualization.

**RNA velocity**—Intronic and exonic matrices were loaded into Python for analysis with scvelo (0.2.2).\textsuperscript{25} Rather than use scvelo’s method to identify dispersed genes, we used the identified differentially expressed genes associated with each trajectory. Likewise, we used the computed aligned principal components from Monocle 3 in scvelo. The full dynamical mode was computed for gene velocities, and this was projected onto the Monocle 3 trajectory. See Figures 1E, 1F, S3B, S3C.

**Gene clustering**—We detail the approach to compute the time to 10\% FC, time to 50\% FC, and time to 95\% FC in the manuscript. We identified clusters by performing k-means clustering for the genes on these three parameters. Our approach to determining the appropriate number of clusters was empirical – we added clusters so long as clusters with new properties emerged, and stopped once adding a cluster resulted in two clusters with near identical properties. This led to an optimal \( k = 5 \). See Figures 1I and 1J.
TF analysis—TF enrichment analysis (Figures 1K, 4D, 4F, 5G) was done using WebGestalt 2019, performing over-representation using the “transcription factor target” database. We started by selecting all TFs with FDR < 0.05. This generally produced a list with significant redundancy; we used different approaches to handle the redundancy depending on the situation. In Figure 1K, we first took all TFs identified as enriched for each cluster (or group of clusters, in the case of LA1 + LA2). We manually aggregated redundant TFs by first labeling TFs into supergroups of interest (for example, MEF2 and RSRFC4 could be combined), and then selecting the term with the highest enrichment. We then selected the top 25 for visualization in the figure by selecting the TFs with the top summed enrichments across all clusters – this inherently picked TFs represented across multiple clusters, though we note that nearly no TFs were identified in only EA1 or LA. For Figure 4D, we were instead interested in top candidates. Thus, we used the affinity propagation method in WebGestalt, which essentially clusters TFs based on their downstream targets. For each cluster, we selected as representative the TF with the highest expression in in vivo CMs (which typically matched the TF selected by WebGestalt). For Figure 5G, we selectively used fold enrichments for our TFs of interest, selecting the term with the highest enrichment ratio. Activity analysis was done in IPA, using the list of differentially expressed genes and the appropriate fold change for each group.

ATAC-seq analysis—Our goal with the ATAC-seq datasets (Figure S4D) was to quantify the percentage of genes in a list of interest with peaks at the promoter-TSS region. In general, we used the peak calling settings from the original manuscripts, typically with q-value <0.05 as a threshold. For Liu et al. and Bertero and Fields et al., the appropriate data was downloaded from GEO as BED or narrowPeaks output files from MACS2. We annotated peaks using HOMER (4.11.1), and subsequently filtered peaks with annotation “promoter-TSS.” For Greenwald and Li et al., data was available as bigWig files. We therefore converted to WIG (bigWigToWig) followed by conversion to BED (wig2bed, 2.4.38). We then called peaks from the BED file using MACS2 (2.2.7.1) using the peak calling settings from the original manuscript as shown on GEO. We then annotated peaks as above with HOMER. The choice of HOMER genome matched the original study, e.g. hg19 for Liu et al. and Greenwald and Li et al., and hg38 for Bertero and Fields et al.

Supplementary Material
Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS
We thank Dr. Deborah Andrew for allowing us to use her group’s COPAS LP-FACS instrument. Conventional FACS experiments were done through the Bloomberg Flow Cytometry and Immunology Core/Cell Sorting Facility with the help of Dr. Hao Zhang. Sequencing experiments were done through Novogene with the help of Sanya Pal, Payton Hovey, Dr. Mary Grantham, and Victor Popoca. Additional assistance was provided by the Johns Hopkins Transcriptomics and Deep Sequencing Core, with the help of Dr. Haiping Hao, Linda Orzolek, Dr. Jasmeet Sethi, and Kelly Laughlin. We additionally thank Dr. Patrick Cahan, Yuqi Tan, Emily Su, and Taibo Li for helpful comments in preparing this manuscript. This work was supported by grants from NICHD/NIH, AHA, and MSCRF.
INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

REFERENCES

1. Kadota S, and Shiba Y (2019). Pluripotent stem cell-derived cardiomyocyte transplantation for heart disease treatment. Curr. Cardiol. Rep 21, 73–19. [PubMed: 31228011]
2. Mummery CL (2018). Perspectives on the use of human induced pluripotent stem cell-derived cardiomyocytes in biomedical research. Stem Cell Rep. 11, 1306–1311.
3. Karbassi E, Fenix A, Marchiano S, Muraoka N, Nakamura K, Yang X, and Murry CE (2020). Cardiomyocyte maturation: advances in knowledge and implications for regenerative medicine. Nat. Rev. Cardiol 17, 341–359. [PubMed: 32015528]
4. Kannan S, and Kwon C (2020). Regulation of cardiomyocyte maturation during critical perinatal window. J. Physiol 598, 2941–2956. [PubMed: 30571853]
5. Kannan S, Farid M, Lin BL, Miyamoto M, and Kwon C (2021). Transcriptomic entropy benchmarks stem cell-derived cardiomyocyte maturation against endogenous tissue at single cell level. PLoS Comput. Biol 17, e1009305. [PubMed: 34534204]
6. Poon E, Keung W, Liang Y, Ramalingam R, Yan B, Zhang S, Chopra A, Moore J, Herren A, Lieu DK, et al. (2015). Proteomic analysis of human pluripotent stem cell-derived, fetal, and adult ventricular cardiomyocytes reveals pathways crucial for cardiac metabolism and maturation. Circ. Cardiovasc. Genet 8, 427–436. [PubMed: 25759434]
7. Cai W, Zhang J, de Lange WJ, Gregorich ZR, Karp H, Farrell ET, Mitchell SD, Tucholski T, Lin Z, Biermann M, et al. (2019). An unbiased proteomics method to assess the maturation of human pluripotent stem cell-derived cardiomyocytes. Circ. Res 125, 936–953. [PubMed: 31573406]
8. Birket MJ, and Mummery CL (2015). Pluripotent stem cell derived cardiovascular progenitors - a developmental perspective. Dev. Biol 400, 169–179. [PubMed: 25624264]
9. Fujita J, Tohyama S, Kishino Y, Okada M, and Morita Y (2019). Concise review: genetic and epigenetic regulation of cardiac differentiation from human pluripotent stem cells. Stem Cell. 37, 992–1002.
10. Uosaki H, Cahan P, Lee DI, Wang S, Miyamoto M, Fernandez L, Kass DA, and Kwon C (2015). Transcriptional landscape of cardiomyocyte maturation. Cell Rep. 13, 1705–1716. [PubMed: 26586429]
11. Murphy SA, Miyamoto M, Kervadec A, Kannan S, Tampakakis E, Kambhampati S, Lin BL, Paek S, Andersen P, Lee DI, et al. (2021). PGC1/PPAR drive cardiomyocyte maturation at single cell level via YAP1 and SF3B2. Nat. Commun 12, 1648. [PubMed: 33712605]
12. Feric NT, and Radisic M (2016). Maturing human pluripotent stem cell-derived cardiomyocytes in human engineered cardiac tissues. Adv. Drug Deliv. Rev 96, 110–134. [PubMed: 25956564]
13. Zhu R, Blazeksi A, Poon E, Costa KD, Tung L, and Boheler KR (2014). Physical developmental cues for the maturation of human pluripotent stem cell-derived cardiomyocytes. Stem Cell Res. Ther 5, 117. [PubMed: 25688759]
14. Nguyen AH, Marsh P, Schmiess-Heine L, Burke PJ, Lee A, Lee J, and Cao H (2019). Cardiac tissue engineering: state-of-the-art methods and outlook. J. Biol. Eng 13. 57–21. [PubMed: 31297148]
15. Scuderi GJ, and Butcher J (2017). Naturally engineered maturation of cardiomyocytes. Front. Cell Dev. Biol 5, 50–28. [PubMed: 28529939]
16. Ahmed RE, Anzai T, Chanthra N, and Uosaki H (2020). A brief review of current maturation methods for human induced pluripotent stem cells-derived cardiomyocytes. Front. Cell Dev. Biol 8, 178. [PubMed: 32266260]
17. Gintant G, Burridge P, Gepstein L, Harding S, Herron T, Hong C, Jalife J, and Wu JC (2019). Use of human induced pluripotent stem cell-derived cardiomyocytes in preclinical cancer drug cardiotoxicity testing: a scientific statement from the American heart association. Circ. Res 125, e75–e92. [PubMed: 31533542]
18. Ovics P, Regev D, Baskin P, Davidor M, Shemer Y, Neeman S, Ben-Haim Y, and Binah O (2020). Drug development and the use of induced pluripotent stem cell-derived cardiomyocytes for disease modeling and drug toxicity screening. Int. J. Mol. Sci 21, 7320–7342. [PubMed: 33023024]

19. Kannan S, Miyamoto M, Lin BL, Zhu R, Murphy S, Kass DA, Andersen P, and Kwon C (2019). Large particle fluorescence-activated cell sorting enables high-quality single-cell RNA sequencing and functional analysis of adult cardiomyocytes. Circ. Res 125, 567–569. [PubMed: 31415233]

20. Ackers-johnson M, Tan WLW, Foo RSY, and Foo RS (2018). Following hearts, one cell at a time: recent applications of single-cell RNA sequencing to the understanding of heart disease. Nat. Commun 9, 4434. 10.1038/s41467-018-06894-8. [PubMed: 30375391]

21. Tan Y, and Cahan P (2019). SingleCellNet: a computational tool to classify single cell RNA-seq data across platforms and across species. Cell Syst. 9, 207–213.e2. [PubMed: 31377170]

22. Maroli G, and Braun T (2020). The long and winding road of cardiomyocyte maturation. Cardiovasc. Res 12, 1–15.

23. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502. 10.1038/s41587-020-0591-3. [PubMed: 32747759]

24. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gorham JM, Kathiriya IS, Hinson JT, Homsy J, Gray J, Pu W, et al. (2016). Single-cell resolution of temporal gene expression during heart development. Dev. Cell 39, 480–490. [PubMed: 27840107]

25. Bergen V, Lange M, Peidli S, Wolf FA, and Theis FJ (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol 38, 1408–1414. 10.1038/s41587-020-0591-3. [PubMed: 32747759]

26. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgener H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, et al. (2018). RNA velocity of single cells. Nature 560, 494–498. [PubMed: 30089906]

27. Campbell KR, and Yau C (2017). Inference of switch-like differential expression along single-cell trajectories. Bioinformatics 33, 1241–1242. [PubMed: 28011787]

28. Campbell KR, and Yau C (2019). A descriptive marker gene approach to single-cell pseudotime inference. Bioinformatics 35, 28–35. [PubMed: 29939207]

29. Guo Y, Jardin BD, Zhou P, Sethi I, Akerberg BN, Toepfer CN, Ai Y, Li Y, Ma Q, Guatimosim S, et al. (2018). Hierarchical and stage-specific regulation of cardiomyocyte maturation by serum response factor. Nat. Commun 9, 3837. [PubMed: 30242271]

30. Sakamoto T, Matsuura TR, Wan S, Ryba DM, Kim JU, Won KJ, Lai L, Petucci C, Petrenko N, Musunuru K, et al. (2020). A critical role for estrogen-related receptor signaling in cardiac maturation. Circ. Res 126, 1685–1702. [PubMed: 32212902]

31. Naya FJ, Black BL, Wu H, Bassel-Duby R, Richardson JA, Hill JA, and Olson EN (2002). Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. Nat. Med 8, 1303–1309. [PubMed: 12379849]

32. Desjardins CA, and Naya FJ (2016). The function of the MEF2 family of transcription factors in cardiac development, cardiogenomics, and direct reprogramming. J. Cardiovasc. Dev. Dis 3, 26. [PubMed: 27630998]

33. Kadari A, Mekala S, Wagner N, Malan D, Köth J, Doll K, Stappert L, Eckert D, Peitz M, Matthes J, et al. (2015). Robust generation of cardiomyocytes from human iPSC cells requires precise modulation of BMP and WNT signaling. Stem Cell Rev. Rep 11, 560–569. [PubMed: 25392050]

34. Cao N, Liang H, Huang J, Wang J, Chen Y, Chen Z, and Yang HT (2013). Highly efficient induction and long-term maintenance of multipotent cardiovascular progenitors from human pluripotent stem cells under defined conditions. Cell Res. 23, 1119–1132. [PubMed: 23896987]

35. Kattman SJ, Witty AD, Gagliardi M, Dubois NC, Niapour M, Hotta A, Ellis J, and Keller G (2011). Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell 8, 228–240. [PubMed: 21295278]

36. Haghverdi L, Lun ATL, Morgan MD, and Marioni JC (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol 36, 421–427. [PubMed: 29608177]
37. Andersen P, Tampakakis E, Jimenez DV, Kannan S, Miyamoto M, Shin HK, Saberi A, Murphy S, Sulistio E, Chelke SP, and Kwon C (2018). Precardiac organoids form two heart fields via Bmp/Wnt signaling. Nat. Commun 9, 3140. 10.1038/s41467-018-05604-8. [PubMed: 30087351]

38. Friedman CE, Nguyen Q, Lukowski SW, Helfer A, Chiu HS, Miklas J, Levy S, Suo S, Han JDJ, Osteil P, et al. (2018). Single-cell transcriptomic analysis of cardiac differentiation from human PSCs reveals HOPX-dependent cardiomyocyte maturation. Cell Stem Cell 23, 586–598.e8. [PubMed: 30290179]

39. Churko JM, Garg P, Treutlein B, Venkatasubramanian M, Wu H, Lee J, Wessells QN, Chen SY, Chen WY, Chetal K, et al. (2018). Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. Nat. Commun 9, 4906. [PubMed: 30464173]

40. Gerbin KA, Grancharova T, Donovan-Maiye RM, Hendershott MC, Anderson HG, Brown JM, Chen J, Dinh SQ, Gehring JL, Johnson GR, et al. (2021). Cell states beyond transcriptomics: integrating structural organization and gene expression in hiPSC-derived cardiomyocytes. Cell Syst. 12, 670–687.e10. [PubMed: 34043964]

41. Ruan H, Liao Y, Ren Z, Mao L, Yao F, Yu P, Ye Y, Zhang Z, Li S, Xu H, et al. (2019). Single-cell reconstruction of differentiation trajectory reveals a critical role of ETS1 in human cardiac lineage commitment. BMC Biol. 17. 89–16. [PubMed: 31722692]

42. Liu Q, Jiang C, Xu J, Zhao MT, Van Bortle K, Cheng X, Wang G, Chang HY, Wu JC, and Snyder MP (2017). Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. Circ. Res 121, 376–391. [PubMed: 28663367]

43. Bertero A, Fields PA, Ramani V, Bonora G, Yardimci GG, Reinecke H, Pabon L, Noble WS, Shendure J, and Murry CE (2019). Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. Nat. Commun 10, 1538–1619. [PubMed: 30948719]

44. Greenwald WW, Li H, Benaglio P, Jakubosky D, Matsui H, Schmitt A, Selvaraj S, D’Antonio M, D’Antonio-Chronowska A, Smith EN, and Frazer KA (2019). Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. Nat. Commun 10, 1054–1117. [PubMed: 3083461]

45. Nomura S, Satoh M, Fujita T, Higo T, Sumida T, Ko T, Yamaguchi T, Tobita T, Naito AT, Ito M, et al. (2018). Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure. Nat. Commun. 9, 4435–4517. 10.1038/s41467-018-06639-7. [PubMed: 30375404]

46. Keenan JE, Sullivan H, Ulrich A, Mao L, and Piantadosi CA (2014). Cardiomyocyte-specific ablation of nuclear respiratory factor 1 in the mouse leads to dysregulation of mitochondrial biogenesis, apoptosis, and heart failure. Circ. Res 115, A86.

47. Gawdzik JC, Yue MS, Martin NR, Elemans LMH, Lanham KA, Heideman W, Rezende R, Baker TR, Taylor MR, and Plavicki JS (2018). Sox9B is required in cardiomyocytes for cardiac morphogenesis and function. Sci. Rep 8, 13906. [PubMed: 30224706]

48. Schauer A, Adams V, Poitz DM, Barthel P, Joachim D, Friedrich J, Linke A, and Augstein A (2019). Loss of Sox9 in cardiomyocytes delays the onset of cardiac hypertrophy and fibrosis. Int. J. Cardiol 282, 68–75. [PubMed: 30765281]

49. Sucharov CC, Dockstader K, and McKinsey TA (2008). YY1 protects cardiac myocytes from pathologic hypertrophy by interacting with HDAC5. Mol. Biol. Cell 19, 4141–4153. [PubMed: 18632988]

50. Gregoire S, Li G, Sturzu AC, Schwartz RJ, and Wu SM (2017). YY1 expression is sufficient for the maintenance of cardiac progenitor cell state. Stem Cell. 35, 1913–1923.

51. Windak R, Müuller J, Felley A, Akhmedov A, Wagner EF, Pedrazzini T, Sumara G, and Ricci R (2013). The AP-1 transcription factor c-Jun prevents stress-imposed maladaptive remodeling of the heart. PLoS One 8, e732944.

52. Qiagen (2014). Ingenuity Upstream regulator analysis in IPA. In Qiagen White Paper (Qiagen), pp. 1–10.

53. Branco MA, Cotovio JP, Rodrigues CAV, Vaz SH, Fernandes TG, Moreira LM, Cabral JMS, and Diogo MM (2019). Transcriptomic analysis of 3D cardiac differentiation of human induced...
pluripotent stem cells reveals faster cardiomyocyte maturation compared to 2D culture. Sci. Rep 9, 9229–9313. [PubMed: 31239450]

54. Zhao Y, Rafatian N, Feric NT, Cox BJ, Aschar-Sobbi R, Wang EY, Aggarwal P, Zhang B, Conant G, Ronaldson-Bouchard K, et al. (2019). A platform for generation of chamber-specific cardiac tissues and disease modeling. Cell 176, 913–927.e18. [PubMed: 30686581]

55. Kuppusamy KT, Jones DC, Sperber H, Madan A, Fischer KA, Rodriguez ML, Pabon L, Zhu WZ, Tulloch NL, Yang X, et al. (2015). Let-7 family of microRNA is required for maturation and adult-like metabolism in stem cell-derived cardiomyocytes. Proc. Natl. Acad. Sci. USA 112, E2785–E2794. [PubMed: 25964336]

56. Feyen DAM, McKeithan WL, Bruyneel AAN, Spiering S, Hörmann L, Ulmer B, Zhang H, Briganti F, Schweizer M, Hegyi B, et al. (2020). Metabolic maturation media improve physiological function of human iPSC-derived cardiomyocytes. Cell Rep. 32, 107925. [PubMed: 32697997]

57. Lam YY, Keung W, Chan CH, Geng L, Wong N, Brenière-Letuffe D, Li RA, and Cheung YF. (2020). Single-cell transcriptomics of engineered cardiac tissues from patient-specific induced pluripotent stem cell-derived cardiomyocytes reveals abnormal developmental trajectory and intrinsic contractile defects in hypoplastic right heart syndrome. J. Am. Heart Assoc 9, e016528. [PubMed: 33059525]

58. Giacomelli E, Meraviglia V, Campostrini G, Cochrane A, Cao X, van Helden RWJ, Kotronenberg Garcia A, Mircea M, Kostidis S, Davis RP, et al. (2020). Human-iPSC-Derived cardiac stromal cells enhance maturation in 3D cardiac microtissues and reveal non-cardiomyocyte contributions to heart disease. Cell Stem Cell 26, 862–879.e11. [PubMed: 32459996]

59. Miki K, Deguchi K, Nakashima-Koakutsu M, Lucena-Cacace A, Kondo S, Fujiiwara Y, Hatani T, Sasaki M, Naka Y, Okubo C, et al. (2021). ERG enhances cardiac maturation with T-tubule formation in human iPSC-derived cardiomyocytes. Nat. Commun 12, 3596–3615. [PubMed: 34155205]

60. Wickramasinghe NM, Sachs D, Shewale B, Gonzalez DM, Dhanan-Krishnan P, Torre D, LaMarca E, Raimo S, Dariolli R, Serasinghe MN, et al. (2022). PPARdelta activation induces metabolic and contractile maturation of human pluripotent stem cell-derived cardiomyocytes. Cell Stem Cell 29, 559–576.e7. [PubMed: 35325615]

61. Callaghan NI, Durland L, Chen W, Kuzmanov U, Miranda MZ, Mirzaei Z, Ireland RG, Wang EY, Wagner K, Kim MM, et al. (2022). Advanced physiological maturation of iPSC-derived human cardiomyocytes using an algorithm-directed optimization of defined media components. Preprint at bioRxiv. 10.1101/2022.10.10.507929.

62. Jonker SS, Louey S, Giraud GD, Thornburg KL, and Faber JJ (2015). Timing of cardiomyocyte growth, maturation, and attrition in perinatal sheep. FASEB J 29, 4346–4357. [PubMed: 26139099]

63. Rog-Zielinska EA, Thomson A, Kenyon CJ, Brownstein DG, Moran CM, Szumska D, Michailidou Z, Richardson J, Owen E, Watt A, et al. (2013). Glucocorticoid receptor is required for foetal heart maturation. Hum. Mol. Genet 22, 3269–3282. [PubMed: 23595884]

64. Rog-Zielinska EA, Craig MA, Manning JR, Richardson RV, Gowsans GJ, Dunbar DR, Gharbi K, Kenyon CJ, Holmes MC, Hardie DG, et al. (2015). Glucocorticoids promote structural and functional maturation of foetal cardiomyocytes: a role for PGC-1α. Cell Death Differ. 22, 1106–1116. [PubMed: 25361084]

65. Li M, Iismaa SE, Naqvi N, Nicks A, Husain A, and Graham RM (2014). Thyroid hormone action in postnatal heart development. Stem Cell Res. 13, 582–591. [PubMed: 25087894]

66. Rupert CE, and Coulombe KLM (2017). IGF1 and NRG1 enhance proliferation, metabolic maturity, and the force-frequency response in hESC-derived engineered cardiac tissues. Stem Cells Int. 2017, 7648409. [PubMed: 28951744]

67. Bagnoli JW, Ziegenhain C, Janjic A, Wange LE, Vieth B, Parekh S, Geuder J, Hellmann I, and Enard W (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat. Commun 9, 2937. [PubMed: 30050112]

68. VanDusen NJ, Guo Y, Gu W, and Pu WT (2017). CASAaV: a CRISPR-based platform for rapid dissection of gene function in vivo. Curr. Protoc. Mol. Biol 120, 31.11.1–31.11.14.

Cell Rep. Author manuscript; available in PMC 2023 October 23.
69. Cardoso-Moreira M, Sarropoulos I, Velten B, Mort M, Cooper DN, Huber W, and Kaessmann H (2020). Developmental gene expression differences between humans and mammalian models. Cell Rep. 33, 108308. [PubMed: 33113372]

70. Usashi H, and Taguchi Y. (2016). Comparative gene expression analysis of mouse and human cardiac maturation. Dev. Reprod. Biol 14, 207–215.

71. Leu M, Ehler E, and Perriard JC (2001). Characterisation of postnatal growth of the murine heart. Anat. Embryol 204, 217–224.

72. Cauquil M, Mias C, Guilbeau-Frugier C, Karsenty C, Seguelas MH, Genet G, Renaud-Gabardos E, Prats AC, Pons V, Branchereau M, et al. (2019). Ephrin-B1 blocks adult cardiomyocyte proliferation and heart regeneration. Preprint at bioRxiv. 10.1101/735571.

73. Bassat E, Mutlak YE, Genzeliakh A, Shadrin IY, Baruch Umansky K, Yifa O, Kain D, Rajchman D, Leach J, Riaov Bassat D, et al. (2017). The extracellular matrix protein agrin promotes heart regeneration in mice. Nature 547, 179–184. [PubMed: 28581497]

74. Jabart E, Molho J, Sin K, Stansfield B, Kazmouz SG, Ventro D, Gardner K, Wu JC, and Churko JM (2020). Single-cell protein expression of hiPSC-derived cardiomyocytes using Single-Cell Westerns. J. Mol. Cell. Cardiol 149, 115–122. [PubMed: 33010256]

75. Melsted P, Booeshaghi AS, Liu L, Gao F, Lu L, Min KHJ, da Veiga Beltrame E, Hjörleifsson KE, Gehring J, and Pachter L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat. Biotechnol 39, 813–818. [PubMed: 33795888]

76. Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550–621. [PubMed: 25516281]

77. Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, Darmanis S, Batson J, Botvinnik O, Chen MB, et al. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 562, 367–372. [PubMed: 30283141]

78. Mi H, Muruganujan A, Casagrande JT, and Thomas PD (2013). Large-scale gene function analysis with the panther classification system. Nat. Protoc 8, 1551–1566. [PubMed: 23868073]

79. Supek F, Bošnjak M, Škunca N, and Šmuc T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 6, e21800. [PubMed: 21789182]

80. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137. 10.1186/gb-2008-9-9-r137. [PubMed: 18798982]

81. Feng J, Liu T, Qin B, Zhang Y, and Liu XS (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc 7, 1728–1740. [PubMed: 22936215]

82. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]

83. Nep S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920. [PubMed: 22576172]
Highlights

- Endogenous cardiomyocytes undergo a critical perinatal window of maturation
- Stem cell-derived cardiomyocytes are arrested at the onset of the perinatal phase
- PSC-CMs show poor expression of maturation-related transcription factors
- Cell engineering methods only partially activate dysregulated transcription factors
Figure 1. scRNA-seq data reconstruct a developmental trajectory of *in vivo* CM maturation
(A) Schematic of the experimental workflow for generating *in vivo* CM libraries.
(B) Trajectory inferred by Monocle 3, labeled by time point. Arrow indicates direction of pseudotime.
(C) Pseudotime scores per time point for the inferred trajectory. We designated pseudotime intervals as embryonic/neonatal, perinatal, and adult, respectively.
(D) Percentage of the total differentially expressed genes that become differentially expressed within a pseudotime interval. We subdivided pseudotime into binned subsets of cells and identified differentially expressed genes for each subset.
(E) RNA velocity stream plot projected onto inferred trajectory, labeled by time point.
(F) RNA velocity length across Monocle 3-inferred pseudotimes.
(G) Schematic of the workflow for computing gene dynamics. We defined the maturation FC as the maximum FC in the pseudotime interval [0, 42], and computed the time to achieve a designated percentage of this FC.
(H) Average gene dynamics for each gene cluster, split into upregulated and downregulated genes for each cluster. Clusters were determined based on gene dynamics parameters.
(I) Gene dynamics parameters (time to 10%, 50%, and 95% FC) for each cluster. EA, early activation; LA, late activation, with groupings based on the time to 10% FC.
(J) Number of upregulated and downregulated genes in each identified cluster.
(K) TFs whose downstream targets are enriched in each gene cluster. As LA2 has relatively few genes, we combined LA1 and LA2 and performed enrichment.
Figure 2. Trajectory reconstruction enables direct comparison of *in vivo* and *in vitro* CM maturation

(A) Schematic of experimental workflow for generating PSC-CM libraries.

(B) Myh6-mG and Tnnt2 immunofluorescence of PSC-CMs. Right image shows a one sample PSC-CM, with clearly visible sarcomeric structure.

(C) Trajectory inferred by Monocle 3, labeled by time point. Arrow indicates the direction of pseudotime.

(D) Pseudotime scores per time point for the inferred trajectory.

(E) Gene Ontology analysis of genes differentially expressed across the PSC-CM trajectory.
(F) Combined in vivo CM and PSC-CM trajectories inferred by Monocle 3, labeled by time point.

(G) Combined in vivo CM and PSC-CM trajectories with group differences explicitly treated as a batch effect by mnnCorrect, inferred by Monocle 3 and labeled by time point.

(H) Pseudotime scores per time point for the combined trajectory in (G).

(I) Combined in vivo CM and PSC-CM trajectories with group differences implicitly treated as a batch effect by mnnCorrect, inferred by Monocle 3 and labeled by time point.

(J) Pseudotime scores per time point for the combined trajectory in (I).

(K) Transcriptomic entropy for in vivo CMs and PSC-CMs across time points.
Figure 3. PSC-CM maturation shows both absolute and relative gene expression differences compared with endogenous maturation

(A) Number of in vivo maturation genes that show global expression differences when comparing all in vivo CMs with PSC-CMs (top) or early-stage CMs (e.g., in pseudotime interval [0, 15]) with PSC-CMs.

(B) Heatmap of candidate sarcomeric, mitochondrial, and ribosomal genes showing global expression differences between in vivo and PSC-CMs. Pseudotime units are in increments of 5, based on the individual trajectories for each group (e.g., Figure 1B for in vivo CMs and 2C for PSC-CMs).

(C) PSC-CM maturation gene expression.

(D) In Vivo CM Pseudotime [0.20] vs In Vivo CM Pseudotime [0.40].

(E) Upregulated Diff. Exp. Genes vs Downregulated Diff. Exp. Genes.

(F) Mouse PSC-CMs.

(G) Human PSC-CMs.
(C) Histogram of scaled PSC-CM expression levels of \textit{in vivo} maturation genes. We scaled expression levels by setting 0 as the lowest level across \textit{in vivo} pseudotime and 1 as the higher level across \textit{in vivo} pseudotime.

(D) FC of maturation genes \textit{in vitro} vs. \textit{in vivo}, compared for embryonic/neonatal time points (e.g. [0, 20], left) and perinatal time points (e.g. [0, 40], right). Points are labeled by whether the gene is also differentially expressed in PSC-CMs.

(E) Venn diagrams of differentially expressed genes across pseudotimes \textit{in vivo} and \textit{in vitro}.

(F) Percentage of maturation genes in each pseudotime subset (as in Figure 1D) correctly differentially regulated in mouse PSC-CMs.

(G) Percentage of genes in each identified gene cluster correctly differentially regulated in mouse PSC-CMs.

(H) Percentage of maturation genes in each pseudotime subset (as in Figure 1D) correctly differentially regulated in human PSC-CMs for four studies.

(I) Percentage of genes in each identified gene cluster correctly differentially regulated in human PSC-CMs for four studies.
Figure 4. Perinatal maturation programs are dysregulated in PSC-CMs

(A) Workflow for identifying dysregulated genes in each PSC-CM dataset.
(B) Percentage of maturation genes in each pseudotime subset (as in Figure 1D) dysregulated in PSC-CMs.
(C) Percentage of genes in each identified gene cluster dysregulated in PSC-CMs.
(D) TFs whose downstream targets are enriched in the consensus dysregulated gene list.
(E) Scaled expression of identified dysregulated TFs in vivo and in PSC-CMs. Pseudotime units are in increments of 5, based on the individual trajectories for each group (e.g., Figures 1B for in vivo CMs and 2C for PSC-CMs). The corresponding protein-gene names are AP1FJ (Jun), NRF2 (Nfe2l2), and ERR1 (Esrra).
(F) IPA activity scores for identified dysregulated TFs in vivo and in PSC-CMs. Note that SOX9 was not included, as an activity score was not assigned by IPA.
Figure 5. Perturbation methods to engineer PSC-CM maturation activate some, but not all, of the dysregulated maturation TFs

(A) Characteristics of studies and perturbations analyzed.

(B) Number of maturation genes showing differential change in the same direction following perturbation as *in vivo* maturation. Only genes differentially regulated in the same direction were analyzed in further steps.

(C) Percentage of genes differentially regulated in each study in same direction as *in vivo* CMs that fell into either the “correctly differentially regulated” or “dysregulated” categories in unperturbed PSC-CMs.
(D) Percentage of maturation genes in each pseudotime subset (as in Figure 1D) correctly differentially regulated for each perturbation method.

(E) Percentage of genes in each identified gene cluster correctly differentially regulated for each perturbation method.

(F) Plot of which dysregulated maturation TFs are upregulated by each perturbation method.

(G) Heatmap of dysregulated maturation TFs whose downstream targets are enriched in the correctly differentially expressed gene lists for each perturbation method.
## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Chemicals, peptides, and recombinant proteins | | |
| NaCl | Sigma Aldrich | S9888, CAS Number:7647-14-5 |
| KCl | Sigma Aldrich | P3911, CAS Number:7447-40-7 |
| NaH₂PO₄ | Sigma Aldrich | S0751, CAS Number:7558-80-7 |
| NaHCO₃ | Sigma Aldrich | S6014, CAS Number:144-55-8 |
| Glucose | Sigma Aldrich | G8270, CAS Number:50-99-7 |
| 2,3-Butanedione monoxime | Sigma Aldrich | B0753, CAS Number:57-71-6 |
| Taurine | Sigma Aldrich | T0625, CAS Number:107-35-7 |
| MgCl₂ | Sigma Aldrich | M8266, CAS Number:7786-30-3 |
| Collagenase Type II | Worthington | CLS-2 |
| Protease | Sigma Aldrich | P5147, CAS Number:9036-06-0 |
| 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N’-(2-ethanesulfonic acid) | Sigma Aldrich | H3375, CAS Number:7365-45-9 |
| TrypLE™ Select Enzyme (1X), no phenol red | Thermo Fisher | 12563011 |
| Gibco™ IMDM | Thermo Fisher | 12440053 |
| Gibco™ Ham’s F-12 Nutrient Mix | Thermo Fisher | 11765054 |
| Gibco™ N-2 Supplement | Thermo Fisher | 17502001 |
| Gibco™ B-27™ Supplement, minus vitamin A | Thermo Fisher | 12587010 |
| Bovine serum albumin | Sigma Aldrich | A9418, CAS Number:9048-46-8 |
| Gibco™ L-Glutamine (200 mM) | Thermo Fischer | 25030081 |
| Gibco™ Penicillin-Streptomycin | Thermo Fisher | 15140148 |
| Ascorbic acid | Sigma Aldrich | AX1775, CAS Number:50-81-7 |
| 1-thioglycerol | Sigma Aldrich | M6145, CAS Number:96-27-5 |
| CHIR99021 | Selleckchem | S1263 |
| BMP4 | R&D Systems | 5020-BP |
| XAV939 | Sigma Aldrich | X3004, CAS Number:284028-89-3 |
| Lactate | Sigma Aldrich | L7022, CAS Number:867-56-1 |
| ClearSort Sheath Fluid | Sony | Lot 1218L345 |
| Critical commercial assays | | |
| COPAS SELECT LP-FACS Platform | Union Biometrica | N/A |
| MoFlo Legacy | MoFlo | N/A |
| MoFlo | XDP | N/A |
| NovaSeq S4 | Ilumina | N/A |
| Deposited data | | |
| scRNA-seq data (including raw reads + mapped non-transformed counts) | GEO | GSE164591 |

*Cell Rep. Author manuscript; available in PMC 2023 October 23.*
| REAGENT or RESOURCE                              | SOURCE          | IDENTIFIER                             |
|------------------------------------------------|-----------------|----------------------------------------|
| Processed data (including R workspaces and other files) | Synapse         | https://www.synapse.org/#!Synapse:syn23667436/files/ |
| Reproducible code (in R and Python)              | Github          | https://github.com/skannan4/cm-dysregulation Zenodo https://doi.org/10.5281/zenodo.7703280 |
| Branco et al. (2019)\(^53\)                     |                 | GSE116574                               |
| Zhao et al. (2019)\(^54\)                       |                 | GSE114976                               |
| Kuppusamy et al. (2015)\(^55\)                  |                 | GSE62913                                |
| Feyen, McKeithan, and Bruynell et al. (2020)\(^56\) |                 | GSE151279                               |
| Lam et al. (2020)\(^57\)                        |                 | GSE157157                               |
| Giacomelli, Meraviglia, and Campostini et al. (2020)\(^58\) |                 | GSE147694                               |
| Wickramasinghe et al. (2022)\(^60\)             |                 | GSE160987                               |
| Callaghan et al. (2022)\(^61\)                  |                 | GSE214617                               |
| Miki, Deguchi, and Nakanishi-Koakutsu et al. (2021)\(^59\) |                 | GSE135319                               |
| Experimental models: Cell lines                 |                 | N/A                                    |
| aMHC x mTmG mouse embryonic stem cells (labeled lines 2, 3, 4, and 6) | This paper | N/A                                    |
| Experimental models: Organisms/strains          |                 | N/A                                    |
| Mouse: aMHC-Cre: B6.FVB-Tg (Myh6-cre)2182 Md/J mice | Jackson Laboratory | Stock No. 011038                      |
| Mouse: mTmG: B6.129(Cg)-Gt (ROSA)26Stv1a16Cd1Tc-tdTomato,–EGFP,LacZ/J | Jackson Laboratory | Stock No. 007676)                     |
| Software and algorithms                         |                 | N/A                                    |
| kallistobustools (0.46.2)                        | Melsted et al.\(^23\) | N/A                                    |
| singleCellNet (0.1.0)                           | Tan and Cahan\(^21\) | N/A                                    |
| Monocle 3 (0.2.3.3)                             | Cao et al.\(^23\) | N/A                                    |
| DESeq2 (1.26.0)                                 | Love et al.\(^26\) | N/A                                    |
| scvelo (0.2.2)                                  | Bergen et al.\(^25\) | N/A                                    |
| Other                                          |                 | N/A                                    |
| mcSCRB-seq protocol                             | protocols.io    | https://doi.org/10.17504/protocols.io.p9kdr4w |

Cell Rep. Author manuscript; available in PMC 2023 October 23.